

"BABEȘ-BOLYAI" UNIVERSITY CLUJ-NAPOCA
FACULTY OF CHEMISTRY AND CHEMICAL ENGINEERING

DOCTORAL THESIS

SUMMARY

LAVINIA LORENA PRUTEANU

DOCTORAL SUPERVISOR
PROF. DR. LORENTZ JÄNTSCHI

2018

"BABEȘ-BOLYAI" UNIVERSITY CLUJ-NAPOCA
FACULTY OF CHEMISTRY AND CHEMICAL ENGINEERING

**MOLECULAR SIMILARITY APPLIED FOR THE INVESTIGATION OF CHEMICAL
PROPERTIES AND BIOLOGICAL ACTIVITIES**

LAVINIA LORENA PRUTEANU

DOCTORAL SUPERVISOR

PROF. DR. LORENTZ JÄNTSCHI

2018

TABLE OF CONTENTS

GENERAL INTRODUCTION	1
LITERATURE SURVEY	4
CHAPTER 1. DESCRIBING MOLECULAR STRUCTURE	5
1.1. Molecular geometry optimization	7
1.2. Symmetry and molecular properties	8
1.3. Chemical structures graphs	10
CHAPTER 2. CORRELATION OF MOLECULAR STRUCTURE WITH PROPERTIES AND BIOLOGICAL ACTIVITIES	17
2.1. Clustering	17
2.2. Similarity measures	18
2.3. Matrix similarity calculation	24
2.4. Molecular descriptors and fingerprints	31
2.5. Quantitative structure-activity/property relationships	38
2.6. Machine learning	40
CHAPTER 3. MOLECULAR EXPRESSION AT TRANSCRIPTOMIC LEVEL	43
3.1. Methods of gene expression evaluation - Microarray vs RNA-Seq	44
3.2. Molecular analysis	48
3.2.1. Differentially expressed genes	49
3.2.2. Signature matching	51
3.2.3. Protein-protein interaction network	52
3.2.4. Co-expression network	53
3.3. Similarity applicability at gene expression level	54

PERSONAL CONTRIBUTIONS	56
CHAPTER 4. RESEARCH FRAMEWORK	57
4.1. Molecular arrangements and their properties	58
4.2. Similarity in classes of compounds	59
4.3. Quantitative structure-property relationships	61
4.4. Arsenates effect <i>in vitro</i>	62
4.5. Investigation of <i>Lactobacillus</i> therapeutic benefits <i>in vitro</i>	64
CHAPTER 5. HYDRATION MODELS OF MONOVALENT IONS	66
5.1. Materials and methods	66
5.1.1. Geometry optimization	66
5.1.2. Ion-water clusters	67
5.1.3. Building congeners cages	67
5.1.4. Symmetry and molecular stability	69
5.2. Results and discussions	70
5.2.1. Stability of ion-water clusters	71
5.2.2. Molecular symmetry	73
5.2.3. Electronegativity	74
5.3. Conclusions	80
CHAPTER 6. MOLECULAR SIMILARITY IN CLASSES OF ANTI- INFLAMMATORIES	81
6.1. Materials and methods	81
6.1.1. Data collection	81
6.1.2. Scaffold based similarity	84
6.1.3. Principal component analysis and data screening	89
6.2. Results and discussions	89
6.2.1. Generating the phylogeny within species	90
6.2.2. Similarity clusters	99
6.2.3. Self organising map	104

6.2.4. Principal component analysis	106
6.3. Conclusions	109
CHAPTER 7. QUANTITATIVE STRUCTURE-PROPERTY RELATIONSHIP ON STEROIDS	110
7.1. Materials and methods	110
7.1.1. Geometry optimization	110
7.1.2. Building the hypermolecule	111
7.1.3. Topological descriptors	111
7.1.4. Significant atom positions	113
7.2. Results and discussions	114
7.2.1. Regression model based on significant atom positions	114
7.2.2. Leave-one-out validation	116
7.2.3. Training vs. test validation	116
7.3. Conclusions	118
CHAPTER 8. UNDERSTANDING THE EFFECT OF ARSENATES TREATMENT ON BREAST CANCER CELL LINES USING GENE EXPRESSION ANALYSIS	119
8.1. Materials and methods	119
8.1.1. Cell culture and the treatment	119
8.1.2. Microarray assay	120
8.1.3. Analysing microarray data	121
8.2. Results and discussions	124
8.2.1. <i>In vitro</i> matrigel assay	124
8.2.2. Effect of arsenates on the regulation of autophagy and apoptosis	125
8.2.3. Shape of the data based on Pearson correlation	128
8.2.4. Principal component analysis	130
8.2.5. Gene set enrichment analysis	131
8.3. Conclusions	134

CHAPTER 9. UNDERSTANDING THERAPEUTIC EFFECTS OF <i>LACTOBACILLUS</i> ON INTESTINAL CELL LINES USING GENE EXPRESSION ANALYSIS	136
9.1. Materials and methods	136
9.1.1. Cell culture	136
9.1.2. <i>Lactobacilli</i> mixture preparation	136
9.1.3. Bacterial treatment	137
9.1.4. Extraction of total RNA	137
9.1.5. Microarray assay	138
9.1.6. Statistical analysis of microarray data	138
9.1.7. Validation of gene expression data	139
9.2. Results and discussions	140
9.2.1. Microarray screening	140
9.2.2. Functional classification of differentially expressed genes	141
9.2.3. Real-Time PCR validation	146
9.2.4. Pathway analysis	146
9.3. Conclusions	148
GENERAL CONCLUSION AND FUTURE PROSPECTS	150
ANNEXES	154
REFERENCES	167
PUBLICATION LIST	

Abstract

An interdisciplinary context is presenting using concepts from chemistry and biochemistry in order to investigate molecular behavior of chemical structures and their effects at the biological systems level underlying molecular similarity. The aim of the study was to describe and understand the correlation between structural characteristics of inorganic and organic compounds as well as their clustering based on similarity and their effects at cellular level. *In vitro* tested compounds include arsenates and probiotic mixtures, starting from the understanding of molecular behavior in terms of the similarity of inorganic compounds (represented by metal ions) and organic (represented by classes of anti-inflammatory compounds, steroidal derivatives) in water and respectively, *in silico*. The "Literature survey" section is structured on three chapters (Chapter 1. Describing molecular structure, Chapter 2. Correlation of the molecular structure with properties and biological activities, Chapter 3. Molecular expression at transcriptomic level). A series of approaches presented in the "Personal contributions" section structured in five chapters were derived around the concepts presented in the previous section. Understanding the molecular behavior of inorganic compounds in water (Chapter 5), of organic compounds *in silico* (Chapter 6) as well as the relationships between structural properties and biological activity (Chapter 7) continued with the characterization of the therapeutic effects of tested compounds at the cellular level in Chapters 8 and 9) (Figure 0).

Keywords

molecular similarity, chemical properties, topological indices, molecular descriptors and fingerprints, biological activities, compounds clusters, arsenates, probiotics, gene expression, transcriptomics

MOLECULAR SIMILARITY APPLIED TO THE INVESTIGATION OF CHEMICAL PROPERTIES AND BIOLOGICAL ACTIVITIES

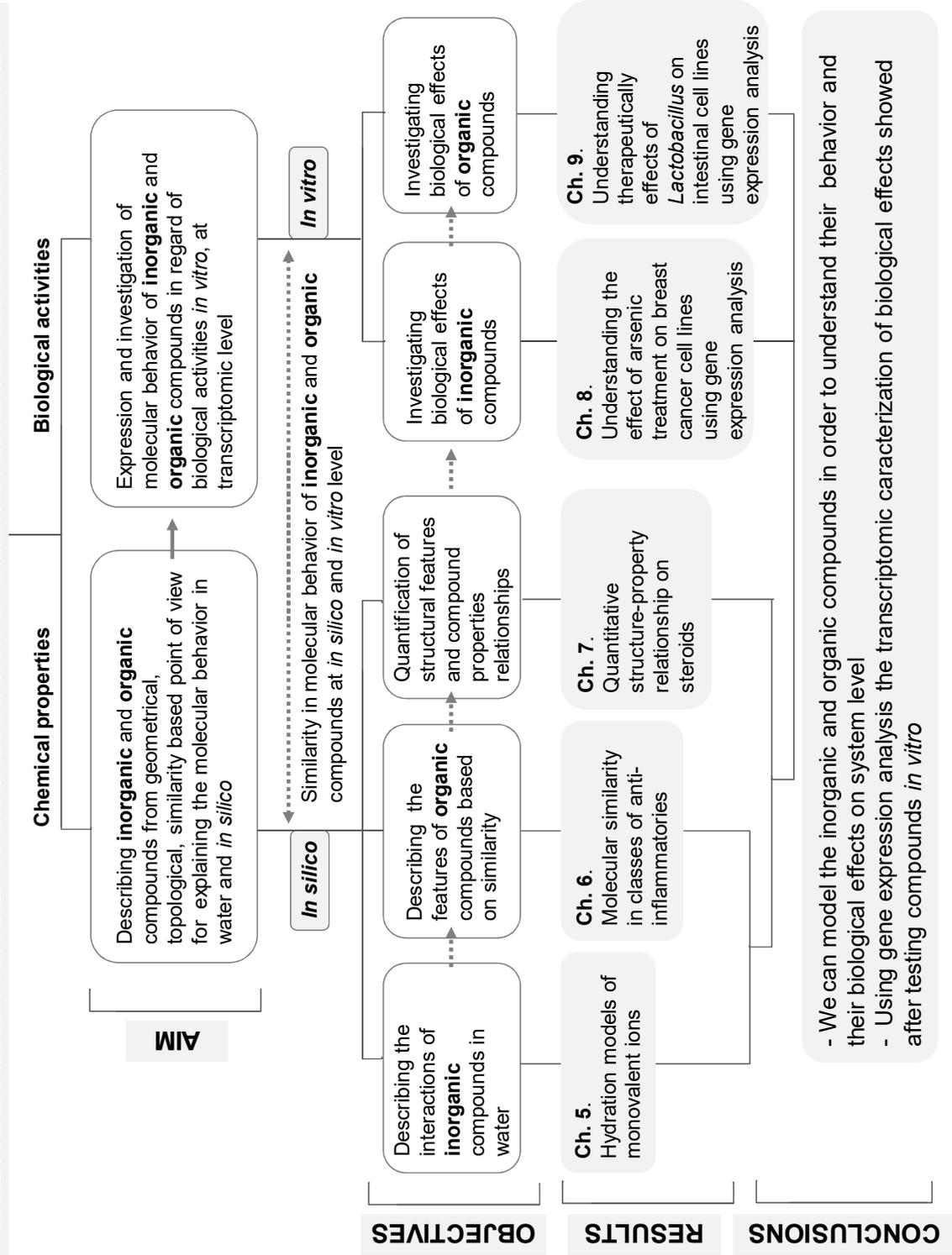


Figure 0. Research framework

CHAPTER 1. DESCRIBING MOLECULAR STRUCTURE

Molecular properties and chemical structure-properties/bioactivity relationships served in understanding the effects on biological systems. Interdisciplinarity is considered a fundamental factor in the research activity carried out.

1.3. Chemical structures graphs

A graph (Figure 1.3.a) is defined as a pair of two sets, a set with a finite number of points (vertices) and a set of discrete pairs of distinct points, while a diagram (Figure 1.3.b) contains a set of finite number of points together with a set of ordered pairs of distinct points. In a multigraph (Figure 1.3.c) we consider two points that can be joined by more than one edge (Diudea et al., 2001).

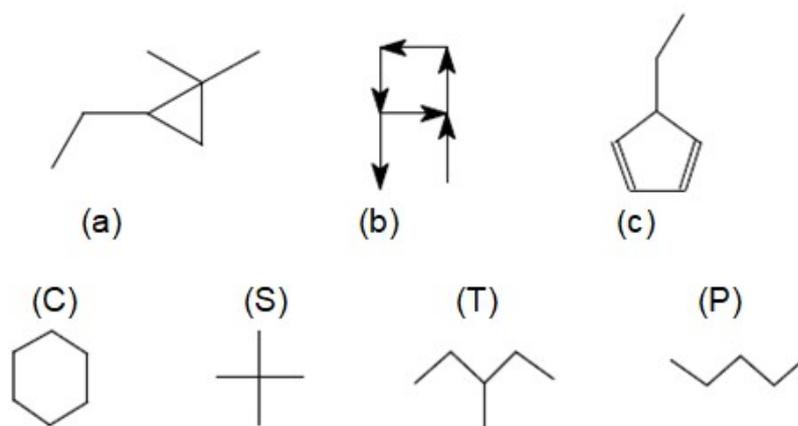
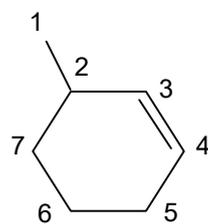


Figure 1.3. Graph representations. Graph (a), digraph (b), multigraph (c); C (cycle): chain that leaves and returns to one and the same vertex; S (star): set of vertices joined by a common line (Sv' , with $v' = v-1$); T (tree): branched structure; P (path): chain (unbranched) - adapted after (Diudea et al., 2001).

The numerical representations of graphs (Figure 1.3) are integrated into algorithms. Representations may be of a functional type for empirical molecular formulas, molecular weights (Harsa et al., 2014), constitutional diagrams (2D), atomic coordinates (Kochev et al., 2003), topological indices, adjacent matrices, distance (Diudea et al., 2007, 2001; Harsa et al., 2014), some of which are given in Table 1.1:

Table 1.1. Exemple de reprezentări funcționale pentru structuri moleculare

Type	Representations
Numbered graph	
Connectivity tables	1,1,2 2,1,1 2,2,1 3,2,4 4,3,2
Liniar anotation	SMILES: <chem>CC(C(=O)C1=CC=CC=C1)O</chem> ROSDAL: 1-2-3-4=5-6-7-2
Adjacency matrix	$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$
Distance matrix	$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & 0 & 1 & 2 & 3 \\ 2 & 1 & 0 & 1 & 2 \\ 3 & 2 & 1 & 0 & 1 \\ 4 & 3 & 2 & 1 & 0 \end{pmatrix}$

The similarity between two entities (atoms, molecules, molecular fragments, even groups of molecules) is defined as a function of their cumulative properties, based on relationships with each other belonging to the same group, this similarity can also be estimated based on topology and molecular geometry.

CHAPTER 2. CORRELATION OF MOLECULAR STRUCTURE WITH PROPERTIES AND BIOLOGICAL ACTIVITIES

Based on specific similar characteristics the molecules can be grouped using clustering algorithms.

2.1. Clustering

Clustering is a proceeding in which different sets of objects/molecules are assigned to certain classes depending on their degree of similarity rendered by spatial or temporal features/properties. Specific empirical calculations are used to identify specific characteristics (Çöltekin et al., 2010; Dodge, 2011).

2.2. Similarity measures

By defining molecular similarity as a measure of the degree of overlapping of two molecules in a given space, we will now recall a series of measures by which we can characterize the similarity between molecules.

Table 2.1. Frequently used similarity measures in binary representations (0;1) for two molecules (A;B)

Measures	Name	Algorithm
Distances	Euclidian	$D_{A,B} = [a + b - 2x]^{1/2}$
	Hamming (Manhattan)	$D_{A,B} = a + b - x$
	Soergel	$D_{A,B} = \frac{1 - x}{a + b - x}$
Similarity coefficients	Tanimoto (Jaccard)	$S_{A,B} = \frac{x}{a + b - x}$
	Dice (Czekanowski)	$S_{A,B} = \frac{2x}{a + b}$
Correlation coefficients	Cosine (Ochiai)	$S_{A,B} = \frac{x}{\sqrt{ab}}$
	Pearson	$S_{A,B} = \frac{ad - bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$

Measures	Name	Algorithm
Legend	A, B - two distinct molecules a - characteristic molecule A b - characteristic molecule B c - distinct characteristic (different then a) molecule A d - distinct characteristic (different then b) molecule B $x = a \cap b$ $D_{A,B}$ = distance between two molecules (A,B) $S_{A,B}$ = similarity between two molecules (A,B)	

To express the similarity of two molecules, the most common measures considered are the Euclidean distances, Hamming (Manhattan), Soergel (Albrecht et al., 2004; Allen et al., 2001; Bero et al., 2017), Tanimoto coefficient (Rogers and Tanimoto, 1960), Dice index (Czekanowski) (Bero et al., 2017; Willett et al., 1998) and Cosine (Ochiai) correlation coefficient (Willett et al., 1998), Pearson (Pearson, 1895) (Table 2.1).

Tanimoto or Jaccard coefficient (Reynolds et al., 1992) is a standard coefficient, dependent on the absence/presence of a characteristic, thus representing a 2D measure (Ma et al., 2011; Willett et al., 1998). In the table below (Table 2.2) the principle of the Tanimoto coefficient is illustrated.

Table 2.2. Principle of the Tanimoto coefficient illustration

a	b	c	d	a	a	d	c	b	A = 3
a	b	b	c	d	a	c	a	a	B = 4
1	1	0	0	0	1	0	0	0	$ A \cap B = 3$
$\text{Tanimoto } (A_a, B_a) = \frac{ A \cap B }{ A + B - A \cap B } = 0.75$									
a, b, c, d - characteristics molecules A, respectively B									

Thus taking two molecules (A, B) and having the representative characteristics (a, b, c, d) of each molecule, it is considered 0 the absence of a certain characteristic (e.g., the absence of an atom in a particular position within a molecule, in comparison with another molecule) and with 1 the presence of that specific characteristic (Table 2.2).

The results obtained applying similarity measures (Figure 2.1) can be used in the establishment of topologies, in the classification and grouping of components belonging to a set of molecules of interest (Diudea et al., 2001).

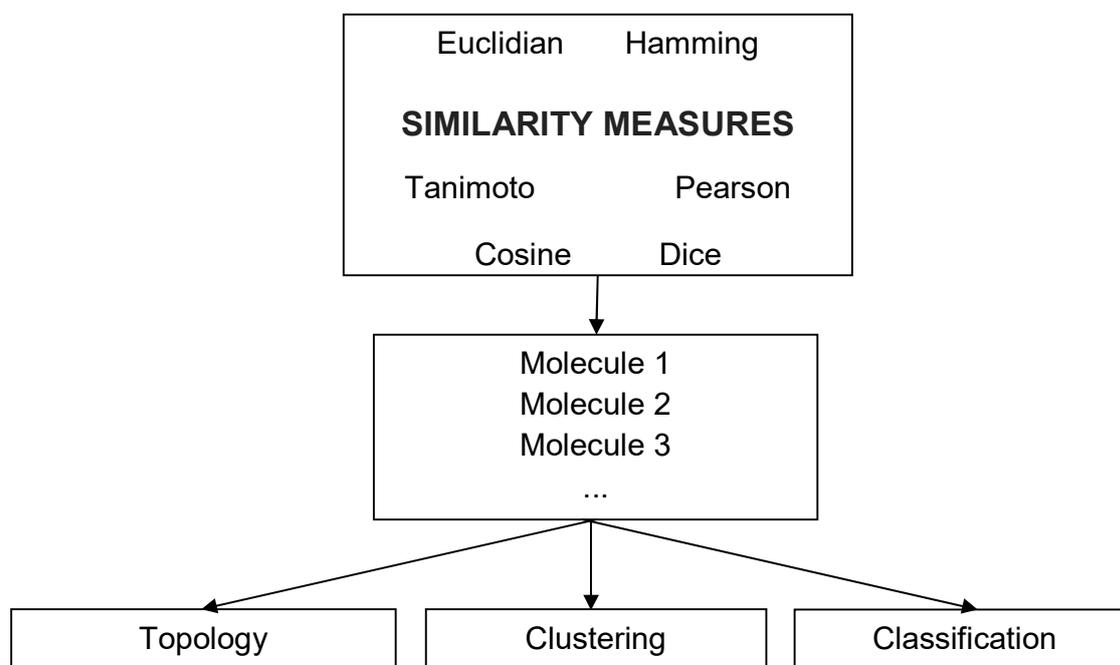


Figure 2.1. Similarity measures can lead to molecular characterization from a topological point of view, to clustering based on similarity and even classification based on certain properties / characteristics.

2.3. Matrix similarity calculation

The matrices can be a measure of the distance between two points/atoms (A, B) (the adjacent matrix) within a molecule located in a space, expressed by a function $d(A, B)$, where $d(A, B) = 0$ if $A = B$ and $d(A, B) = 1$ otherwise. The result of an association of numerical values corresponding to the atomic distances within the molecules takes into account 3 axioms, namely:

- 1) $d(A, B) \geq 0$ și $d(A, B) = 0, A=B$ (reflexivity)
- 2) $d(A, B) = d(B, A)$ (symmetry or commutativity)
- 3) $d(A, B) + d(B, C) \geq d(A, C)$ (subadditivity)

Similarity measures in 2D space allow the topological representation of chemical structures (Table 2.3):

Table 2.5. Distance matrix representation for m1, m2, m3

[Di]	1	2	3	4	5	6	7		[Di]	1	2	3	4	5	6	7	8		[Di]	1	2	3	4	5	6	7	8
1	0	1	1	2	3	2	3		1	0	1	1	2	3	2	3	2		1	0	1	1	2	3	2	3	1
2	1	0	2	1	2	3	4		2	1	0	2	1	2	3	4	3		2	1	0	2	1	2	3	4	2
3	1	2	0	3	2	1	2		3	1	2	0	3	2	1	2	1		3	1	2	0	3	2	1	2	2
4	2	1	3	0	1	2	3		4	2	1	3	0	1	2	3	4		4	2	1	3	0	1	2	3	3
5	3	2	2	1	0	1	2		5	3	2	2	1	0	1	2	3		5	3	2	2	1	0	1	2	4
6	2	3	1	2	1	0	1		6	2	3	1	2	1	0	1	2		6	2	3	1	2	1	0	1	3
7	3	4	2	3	2	1	0		7	3	4	2	3	2	1	0	3		7	3	4	2	3	2	1	0	4
									8	2	3	1	4	3	2	3	0		8	1	2	2	3	4	3	4	0
	m1								m2								m3										

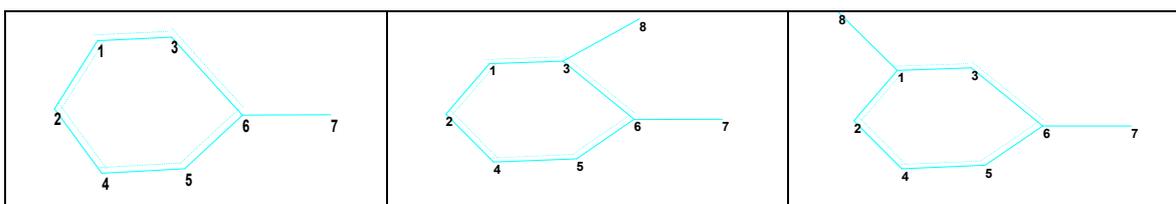
În Table 2.6 și Table 2.7 au fost considerate secvențe de numere și descriptori topologici (aplicabile atât la linii cât și la coloane, rezultând același lucru) care permit analiza de similaritate intramoleculară (în interiorul moleculei) (Table 2.6) și respectiv intermoleculară (între molecule) (Table 2.7):

In Table 2.6 and Table 2.7 there were considered sequence of numbers and topological descriptors (applicable to both lines and columns, resulting in the same) that allowed the intramolecular similarity analysis (inside the molecule) (Table 2.6) and intermolecular (between molecules) (Table 2.7):

Table 2.6. Distance matrix representation for m1, m2, m3

– intramolecular similarity

[Di]	1	2	3	4	5	6	7	0.1.2.3.4	Σ	[Di]	1	2	3	4	5	6	7	8	0.1.2.3.4	Σ	[Di]	1	2	3	4	5	6	7	8	0.1.2.3.4	Σ
1	0	1	1	2	3	2	3	1.2.2.2.0	12	1	0	1	1	2	3	2	3	2	1.2.3.2.0	14	1	0	1	1	2	3	2	3	1	1.3.2.2.0	13
2	1	0	2	1	2	3	4	1.2.2.1.1	13	2	1	0	2	1	2	3	4	3	1.2.2.2.1	16	2	1	0	2	1	2	3	4	2	1.2.3.1.1	15
3	1	2	0	3	2	1	2	1.2.3.1.0	11	3	1	2	0	3	2	1	2	1	1.3.3.1.0	12	3	1	2	0	3	2	1	2	2	1.2.4.1.0	13
4	2	1	3	0	1	2	3	1.2.2.2.0	12	4	2	1	3	0	1	2	3	4	1.2.2.2.1	16	4	2	1	3	0	1	2	3	3	1.2.2.3.0	15
5	3	2	2	1	0	1	2	1.2.3.1.0	11	5	3	2	2	1	0	1	2	3	1.2.3.2.0	14	5	3	2	2	1	0	1	2	4	1.2.3.1.1	15
6	2	3	1	2	1	0	1	1.3.2.1.0	10	6	2	3	1	2	1	0	1	2	1.3.3.1.0	12	6	2	3	1	2	1	0	1	3	1.3.2.2.0	13
7	3	4	2	3	2	1	0	1.1.2.2.1	15	7	3	4	2	3	2	1	0	3	1.1.2.3.1	18	7	3	4	2	3	2	1	0	4	1.1.2.2.2	19
										8	2	3	1	4	3	2	3	0	1.1.2.3.1	18	8	1	2	2	3	4	3	4	0	1.1.2.2.2	19
	m1										m2										m3										



It can be observed that the sequence of numbers "1.2.3.1.0", "1.2.2.2.0", "1.3.2.1.0", "1.2.2.1.1", "1.1.2.2.1" correctly identifies the similarities between atoms - intramolecular similarity.

Not the same can be said about the sum although it can be in this case a similarity indicator as well. The conclusion is that for the analysis of similarity between atoms should be used the sequences of the type "1.2.3.1.0", "1.2.2.2.0", "1.3.2.1.0", "1.2.2.1.1" (Jäntschi, 2000). For intermolecular similarity analysis (Table 2.7) one way to expand is to sum up and another is to keep distinctness.

Table 2.7. Distance matrix representation for m1, m2, m3
– intermolecular similarity

[Di]	1	2	3	4	5	6	7	0.1.2.3.4	Σ
1	0	1	1	2	3	2	3	1.2.2.2.0	12
2	1	0	2	1	2	3	4	1.2.2.1.1	13
3	1	2	0	3	2	1	2	1.2.3.1.0	11
4	2	1	3	0	1	2	3	1.2.2.2.0	12
5	3	2	2	1	0	1	2	1.2.3.1.0	11
6	2	3	1	2	1	0	1	1.3.2.1.0	10
7	3	4	2	3	2	1	0	1.1.2.2.1	15
0.1.2.3.4								0.7.0.0.0.. 0.1.5.1.0.. 0.0.5.2.0.. 5.2.0.0.0	
								7.14.14.10.2	

[Di]	1	2	3	4	5	6	7	8	0.1.2.3.4	Σ
1	0	1	1	2	3	2	3	2	1.2.3.2.0	14
2	1	0	2	1	2	3	4	3	1.2.2.2.1	16
3	1	2	0	3	2	1	2	1	1.3.3.1.0	12
4	2	1	3	0	1	2	3	4	1.2.2.2.1	16
5	3	2	2	1	0	1	2	3	1.2.3.2.0	14
6	2	3	1	2	1	0	1	2	1.3.3.1.0	12
7	3	4	2	3	2	1	0	3	1.1.2.3.1	18
8	2	3	1	4	3	2	3	0	1.1.2.3.1	18
									0.7.0.0.0.. 0.2.4.2.0.. 0.0.4.4.0.. 4.4.0.0	
									8.16.20.16.4	

[Di]	1	2	3	4	5	6	7	8	0.1.2.3.4	Σ
1	0	1	1	2	3	2	3	1	1.3.2.2.0	13
2	1	0	2	1	2	3	4	2	1.2.3.1.1	15
3	1	2	0	3	2	1	2	2	1.2.4.1.0	13
4	2	1	3	0	1	2	3	3	1.2.2.3.0	15
5	3	2	2	1	0	1	2	4	1.2.3.1.1	15
6	2	3	1	2	1	0	1	3	1.3.2.2.0	13
7	3	4	2	3	2	1	0	4	1.1.2.2.2	19
8	1	2	2	3	4	3	4	0	1.1.2.2.2	19
									8.16.20.14.6	

	m1	m2	m3

Thus, considering the matrix (Table 2.7), and computing the Minkowski distance (Minkowski, 1953) between two atoms/two edges (a generalized form of Manhattan distance - see Table 2.1), $d = \sum |p_i - q_i|$, where p_i = number of edges at distance i for

molecule A, q_i = number of edges at distance i for molecule B, d = distance, is obtained (Table 2.7):

$$d(m1,m2) = |7-8| + |14-16| + |14-20| + |10-16| + |2-4| = 1 + 2 + 6 + 6 + 2 = 17$$

$$d(m1,m3) = |7-8| + |14-16| + |14-20| + |10-14| + |2-6| = 1 + 2 + 6 + 4 + 4 = 17$$

$$d(m2,m3) = |8-8| + |16-16| + |20-20| + |16-14| + |4-6| = 4.$$

2.4. Molecular descriptors and fingerprints

Consideration of several molecular descriptors (see Figure 2.3) is very often found in similarity analyzes (Bender et al., 2009), so that by different combinations they result in new descriptors and then their applicability follows correlation methods or main component analyzes.

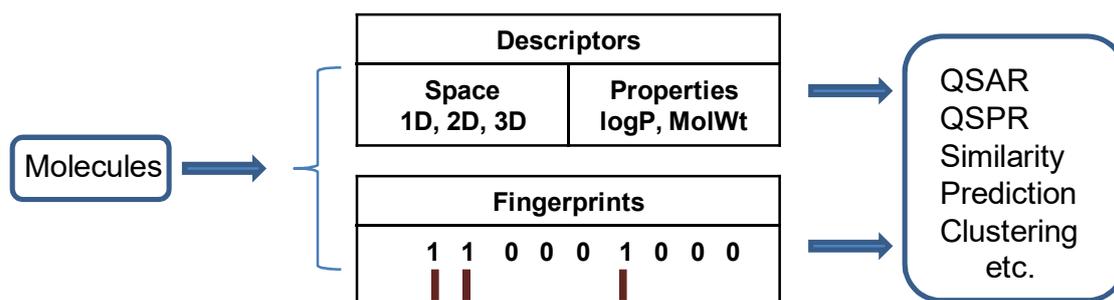


Figure 2.3. Using molecular descriptors and fingerprints in chemistry. The calculation of similarity between two molecules is based on space (1D, 2D, 3D) and molecular properties (lipophilicity - logP, molecular weight - MolWt) and numerical calculations are made using fingerprints as 1100 ... 00. The results allow the initiation of analyzes such as QSAR, QSPR, molecular similarity, prediction, clustering, etc. - figure adapted after (Dong et al., 2015).

The fingerprints (Doucet și Weber, 1996; Rarey și Dixon, 1998) does no represent information regarding to coordinates of a molecule, but these can codify a molecular structure in series of binary values (eg. 01001) (see Figure 2.3).

Three types of topological indices are described in literature as most commonly used, namely:

- Wiener index: integer numerical representations obtained from graphs correlated with some properties (Singh et al., 2008);

- Molecular connectivity indices: in this case, they are represented by real numbers and are held out of graphs with integer numbers correlated with some properties (Diudea et al., 2001);
- Topological indices represented by real numbers derived from graphs with real numbers correlated with certain properties (Bender, 2005).

The molecular scaffolds represent a common method in the search for similarity (Egieyeh et al., 2016; Velkoborsky și Hoksza, 2016). By using a scaffold-like molecule as a model from a set of molecules, a "query" model is created based on the similarity (Willett, 2011).

2.5. Quantitative structure-activity/property relationships

By applying QSAR (Quantitative Structure-Activity Relationship) or QSPR (Quantitative Structure-Property Relationship), it is possible to describe the chemical structure - biological activity, property - biological activity or even chemical structure - physical property relationships which can leads to analyze and collect data both in terms of quantity and qualitatively, and it is then possible to interpret the degree of similarity between certain compounds.

2.6. Machine learning

Machine learning allow the application of computational algorithms (eg Matrices, linear models) in the computer environment (through programming languages and specific computing environments - R in RStudio) (RStudio, 2014) on real sets of experimental data. Among the alternatives that can be used are:

- Combinations of structural chemical descriptors and target sequences can be generated using SVMs (support vector machines) embedded analyzes (Bleakley și Yamanishi, 2009; Jiang et al., 2007; Mishra et al., 2010);
- Measuring similarity (Ding et al., 2013): generation of similarity matrices for pairs of compounds related to their chemical structures occurs. Among the methods that include them, it can be mentioned the Kernel regression (Chen et al., 2015), BLM (bipartite local method) (Bleakley și Yamanishi, 2009), PKM (pairwise kernel method) (Ding et al., 2013), least squares method (Legendre, 1805), Laplacian

(Terfloth, 2003; Thomson et al., 2003), Gaussian interactions (Doucet și Weber, 1996; Jäntschi et al., 2015), Bayesian matrix (Nidhi et al., 2006).

CHAPTER 3. MOLECULAR EXPRESSION AT TRANSCRIPTOMIC LEVEL

The transcriptome is defined as representing the set of all RNA molecules in a cell or a population of cells (Trapnell et al., 2011). These include RNA encoding (RNAc) and non-coding (RNAc), long non-coding RNA (RNAI), small nuclear RNA (RNA), ribosomal RNA (RNA).

The most common type of RNA studied is messenger RNA due to its function of transporting genetic information from DNA required for protein synthesis (Wang et al., 2009). This variety enabled the development of different technologies to understand and predict gene expression in biological systems.

The effects of the compounds are investigated at the level of the biological systems (Bose, 2013) through responses expressed at the genome level, predicting the mechanism of action being possible. The relationship between the biological responses and the mode of action of a compound described at the genetic level can lead to the identification of prognostic or diagnostic biomarkers.

3.1. Methods of gene expression evaluation - Microarray vs RNA-Seq

Microarray techniques allow the measurement of changes in the expression of thousands of genes following their exposure to the action of a particular drug compound (Yu et al., 2006). The principle of operation is based on the attachment of DNA molecules to a solid surface, often called *chip* (Figure 3.2).

Compared to microarray analysis, new techniques such as RNA-Seq, in addition to shortening the DNA extraction and sequencing period, allow for the analysis of very old DNA fragments and the analysis of continuous changes at the level of transcriptoma, allowing direct quantification of gene expression .

The RNA-Seq principle is based on RNA isolation, complementary DNA (cDNA) conversion (Sequences 1, Sequences 2), quality control, sequence library sequencing, and sequencing on a Next Generation Sequencing (NGS) platform.

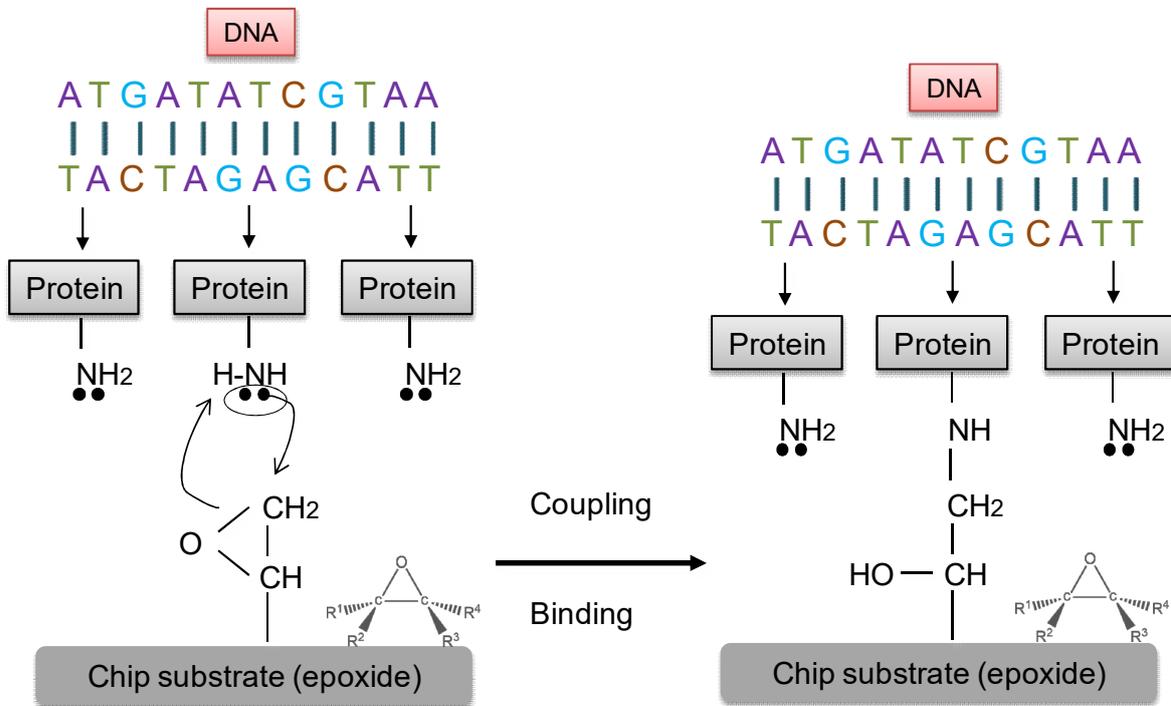


Figure 3.2. The principle of microarray attachment of DNA molecules on the solid surface of the chip by covalent linkages through amino-aliphatic (NH₂) groups - Figure made using the working instrument - ChemBioDraw Ultra 12.0 (Milne, 2010) and adapted after (Stears et al., 2003).

RNA-Seq technique (RNA sequencing) compared to microarray allows for a more detailed analysis from the point of view of:

- gene expression measurement;
- understanding the alternative splicing of genes;
- post-transcriptional modifications.

The generated data can be analyzed to identify new transcription factors, new combinations of alternative genes, and understanding the functionality of the transcriptoma (mRNA, ncRNA, lncRNA) (Kukurba și Montgomery, 2016) (Figure 3.3).

If the microarray is based on the potential for hybridization to samples which are labeled with sequences of the target cDNA, RNA-Seq use advanced techniques for

sequencing, such as Next-Generation Sequencing (NGS) for sequencing RNA by direct sequencing of the strand cDNA (Figure 3.3). In both cases (Bourdon-Lacombe et al., 2015), after obtaining the differentially expressed genes, cellular functionality and pathways expression can be further pursued using the statistical methods and bioinformatics packages that will be mentioned below.

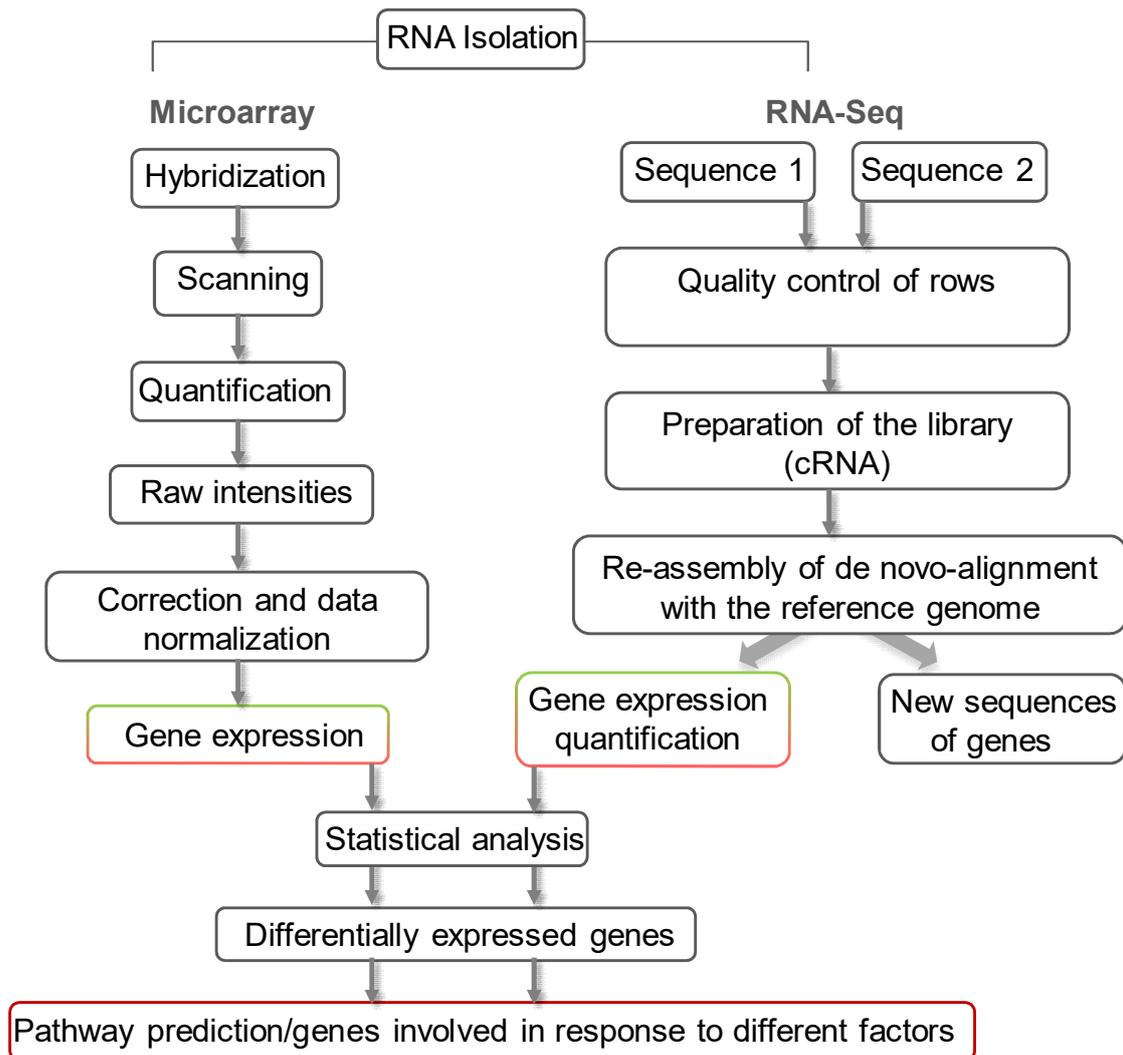


Figure 3.3. Microarray vs RNA-Seq – adapted after (Corney et al., 2013).

3.2. Molecular analysis

For gene expression analysis, a variety of screening methods can be used to identify pathways and genes expressed in response to different factors: determination of

differentially expressed genes (DEGs); signature matching; protein-protein (I-P-P) (network) networks (PPI network); networks of coexpression network (Alexander-Dann et al., 2018) (Figure 3.4).

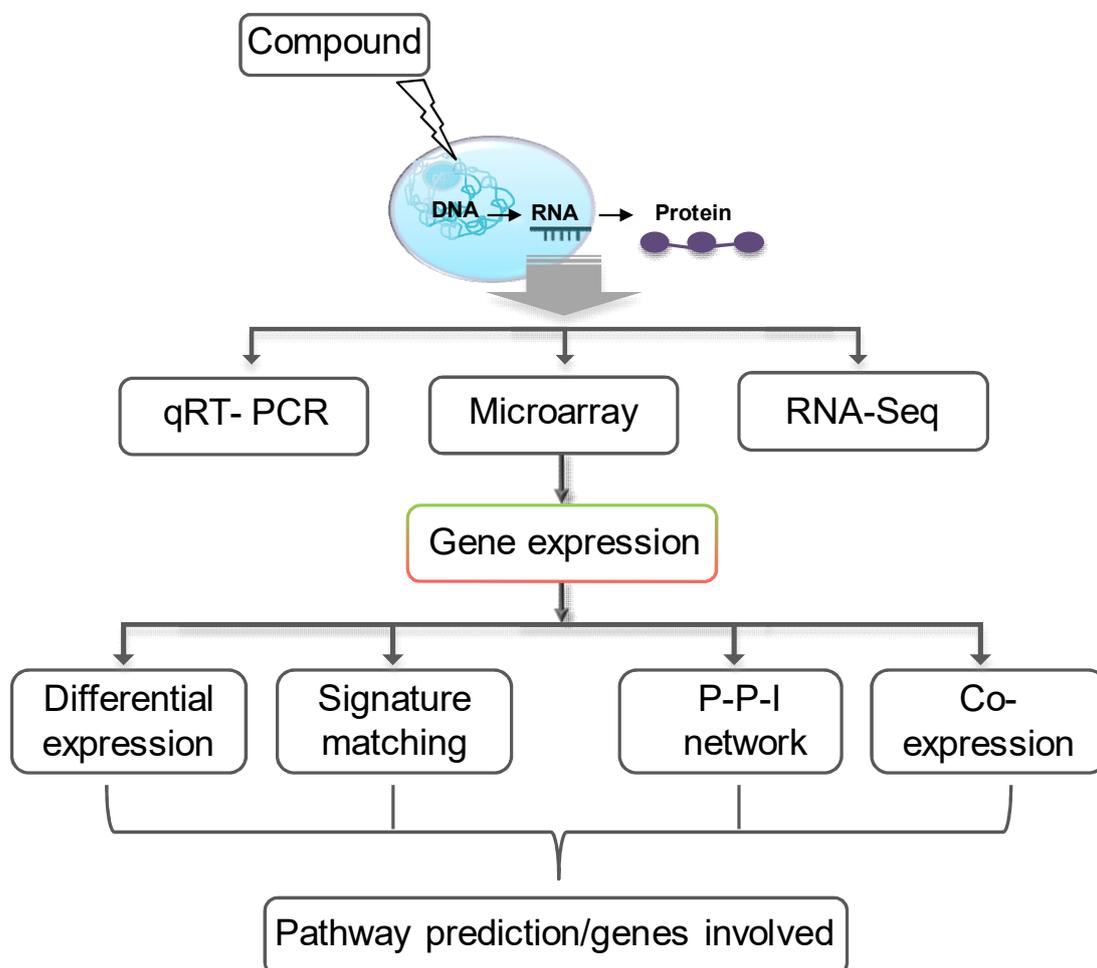


Figure 3.4. Types of analysis of gene expression following cellular exposure to a particular compound - adapted after (Alexander-Dann et al., 2018).

3.2.1. Differentially expressed genes

Differentially expressed genes are useful in the process of identifying specific biomarkers such as tumor biomarkers, toxicity, prognosis or diagnosis (Shi et al., 2008).

An ideal marker is detected prior to traditional pathological symptoms and is characterized by high specificity and sensitivity as well as mechanical relationships of

biological relevance (Rininger et al., 2000). A gene is considered to be differentially expressed if the observed difference between two different conditions is statistically significant (Alexander-Dann et al., 2018; Anjum et al., 2016) Table 3.1.

Tableul 3.1. Analytical measures to determine a DEGs

Method	Principle	Calculation	Utility
Fold change	Calculate the ratio in gene expression between sample and control. It is expressed in the logarithm in the base 2, where it is equal to 0, if there is no difference. Depending on the cut-off, DEGs with values between 0.5 and 2 are determined.	$F_g = \log_2(E_{ga}) - \log_2(E_{gb})$ F - fold change g - gene E - expression a - condition a (ex. treated) b- condition b (ex. untreated)	<ul style="list-style-type: none"> - easy to interpret - small samples - variations occur
Rank product	Comparing gene rows by expression by applying a nonparametric test.	$R_g = \frac{E_{ga}}{E_{gb}}$ R - ratio g - gene E - expression a - condition a (ex. treated) b- condition b (ex. untreated)	<ul style="list-style-type: none"> - comparison of results from different platforms - used in meta-analyzes
t-test, ANOVA	Comparison of the mean value of gene expression in samples. It is based on the null hypothesis where the averages are equal. While the t-test is used to compare two samples, ANOVA allows comparison for multiple samples.	$t = \frac{\text{average}(d)}{\frac{\sigma(d)}{\sqrt{n}}}$ d = $x_a - y_b$ x_a - the mean of the number of genes attributed to condition a y_b - the mean of the number of genes attributed to condition b $\sigma(d)$ variability of genes, σ = standard deviation n - number of tested pairs	<ul style="list-style-type: none"> - high statistical significance - the values are normally distributed and have the same variation
Bayesian	Uses data for prediction of differentiated expression probabilities and standard deviation.	$P(A B) = \frac{P(B A) * P(A)}{P(B)}$ P(A B) - probability of condition A if condition B is present	<ul style="list-style-type: none"> - time consuming - more significant than t-test results

Method	Principle	Calculation	Utility																
		$P(B A)$ - probability of condition B if condition A is present $P(A)$ - probability of priority condition A $P(B)$ - probability of priority condition B																	
Exact test	For comparing gene expression, it uses real expression counting.	$p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!n!}$ <p>where,</p> <table border="1"> <thead> <tr> <th></th> <th>C_A</th> <th>C_B</th> <th>Total</th> </tr> </thead> <tbody> <tr> <td>Gene X</td> <td>a</td> <td>b</td> <td>a+b</td> </tr> <tr> <td>Rest gene</td> <td>c</td> <td>d</td> <td>c+d</td> </tr> <tr> <td>Total</td> <td>a+c</td> <td>b+d</td> <td>n</td> </tr> </tbody> </table> <p> C_A - condition A C_B - condition B </p>		C_A	C_B	Total	Gene X	a	b	a+b	Rest gene	c	d	c+d	Total	a+c	b+d	n	- requires exact copies of mRNA
	C_A	C_B	Total																
Gene X	a	b	a+b																
Rest gene	c	d	c+d																
Total	a+c	b+d	n																

To measure these, a number of statistical methods are considered, such as: fold change (level of gene expression) (Love et al., 2014; Tarca et al., 2006), Rank product (Breitling et al., 2004; Hong et al., 2006), linear methods (ANOVA, t-test) (Trapnell et al., 2013), Bayesian methods (Hardcastle și Kelly, 2010), exact test (Auer și Doerge, 2010; Love et al., 2014), embedded in various computing packages such as *limma* (Tarca et al., 2006), weighted average difference method (Kadota et al., 2008), RankProd (Hong et al., 2006), Cuffdiff 2 (Trapnell et al., 2013), baySeq (Hardcastle și Kelly, 2010), DESeq2 (Love et al., 2014), edgeR (Robinson et al., 2009). The utility and principle of the methods is shown in Table 3.1.

3.2.2. Signature matching

In the methods of Signature matching, DEGs are assessed against a library containing transcriptome profiles in order to predict potential effects (Hettne et al., 2013).

By comparing the changes in gene expression resulting from exposure to certain compounds, their effects on a biological system with their structural properties can be

correlated, separated by what is a molecular similarity analysis. Using also reference gene reference libraries, the similarity of the compounds in the gene expression space can be measured (Lamb et al., 2006).

3.2.3. Protein-protein interaction network

Protein-protein interaction network (PPI network) allows the visualization and characterization of the biological response after the exposure of a particular compound (therapeutic effect or toxic effect - see Figure 3.5) at the cellular level.

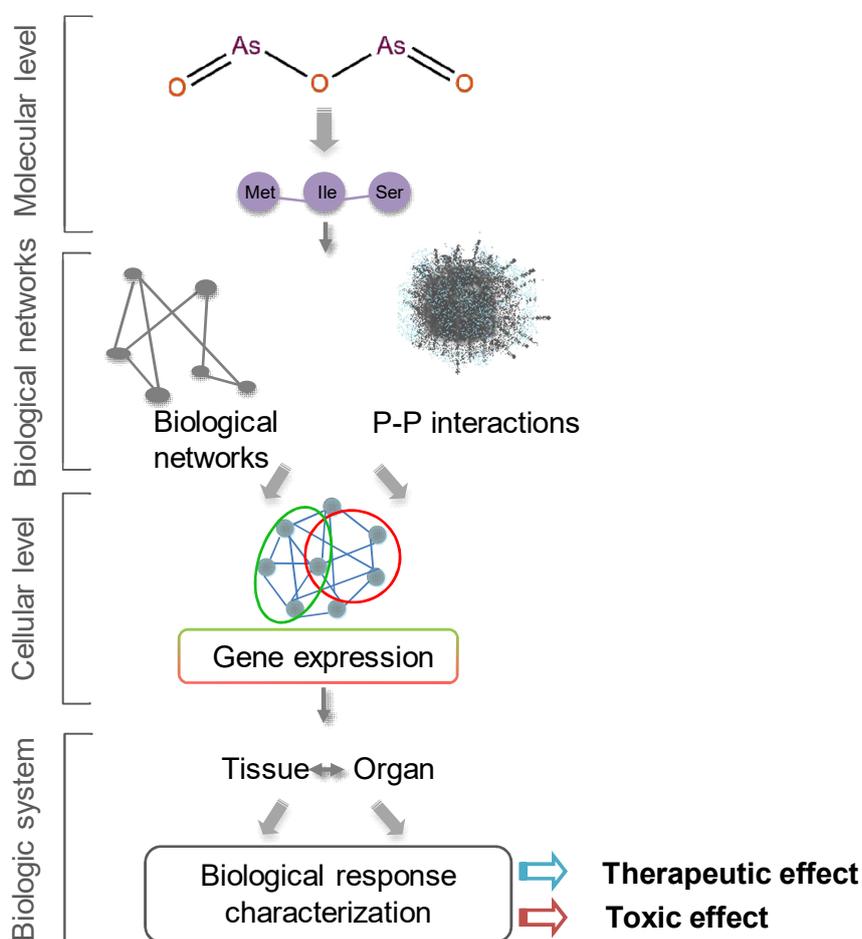


Figure 3.5. Exemplified representation (compound: arsenic trioxide) of the integration of experimental and computational data in biological systems to characterize biological responses through association networks (biological pathways and protein interactions) - Figure adapted after (Orozco et al., 2016).

In Figure 3.5, we can see how arsenic trioxide acts on targets on a molecular level. The biological pathways involved and the interactions between proteins can be analyzed by differentiating expression of the resulting genes. The characterization of the biological response occurs depending on the predictions of the pathways involved.

3.2.4. Co-expression network

Co-expression networks use all of the transcriptom-related measures, which creates a strong correlation between genes and biological effects. Methods applied to date are found in examples such as: Context Likelihood of Relatedness (Taylor et al., 2008), Weighted Gene Co-expression Network Analysis' – WGCNA (Zhang și Horvath, 2005) .

The methods allow the correlation and comparison of the networks of gene expression with a strong impact in predicting the effects of the tested compounds. The similarities and differences of each response for each compound are thus highlighted. The dependence of co-expression methods on determining correlations between gene expressions is associated with the use of a minimum number of replicates.

3.3. Similarity applicability at gene expression level

The transfer of the similarity concept to the gene expression level has been increasingly advanced in recent years. The entire libraries have been created with descriptive information so that the similarity of the compounds can be measured in the expression space of the genes. One of the current approaches refers to Connectivity mapping that can describe the action of compounds at genes level (Sirci et al., 2017).

Through numerical representations, the relationships between the compounds and the expression of the genes can be analyzed. A series of measures to calculate the similarity of gene expression have been developed using graphs as representations

of genes, where the nodes are considered to be genes, and the edges link the genes if there is a relationship between them (Yona et al., 2006).

PERSONAL CONTRIBUTIONS

Molecular similarity has been illustrated by describing the behavior of inorganic compounds in water (Chapter 5), continuing with the description of the characteristics of organic compounds based on similarity (Chapter 6) and the quantification of relations between structural characteristics and properties (Chapter 7), reaching modeling and characterization at the transcriptomic level of the response of the inorganic compounds (arsenates) (Chapter 8), the probiotics (Chapter 9) following the treatment of the breast and intestinal epithelial cell lines *in vitro* (Figure 4.1).

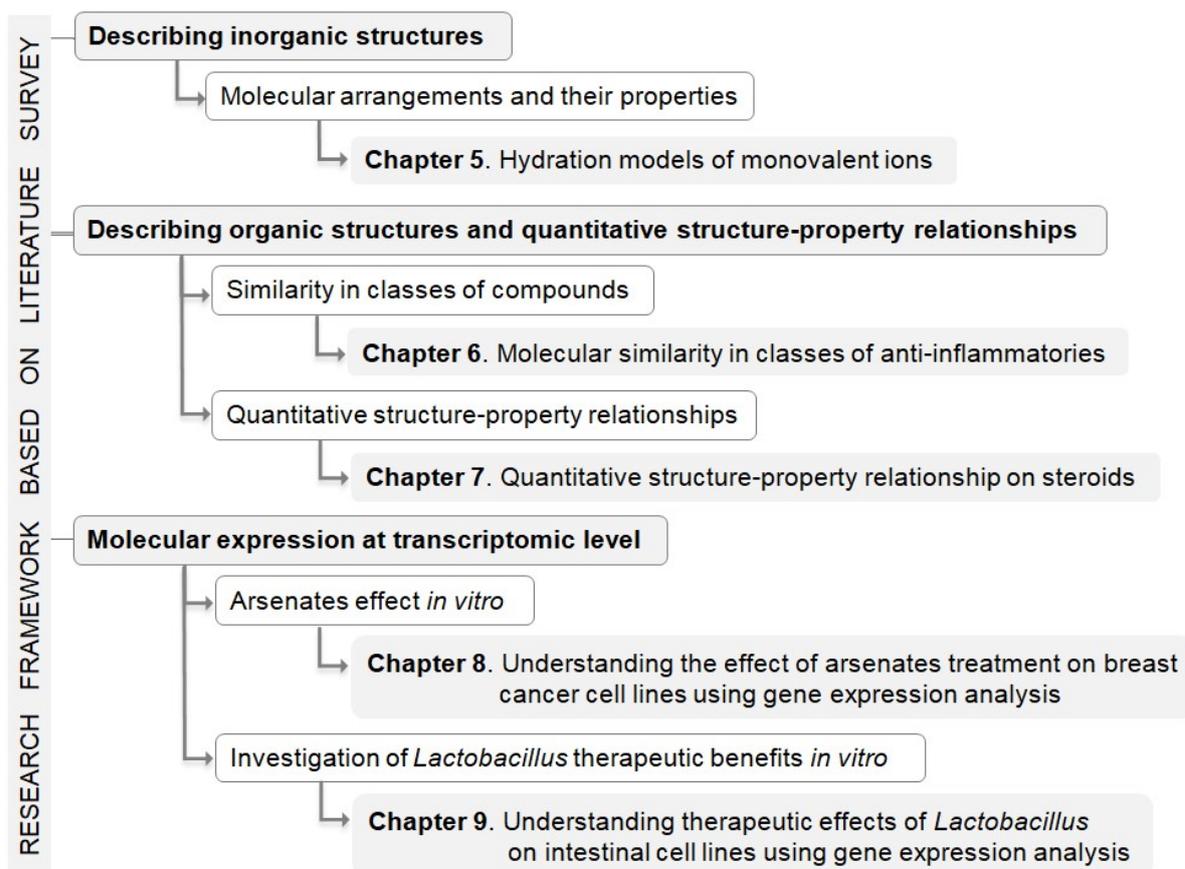


Figure 4.1. Research framework based on literature survey

CHAPTER 4. RESEARCH FRAMEWORK

4.1. Molecular arrangements and their properties

In order to investigate and describe the behavior of inorganic structures in water, the structural characteristics and their properties were observed in the resulting molecular arrangements. In this context, Chapter 5 presents the hydration models obtained on the basis of the interaction modeling of six monovalent ions (NH_4^+ , F^- , Cl^- , Li^+ , Na^+ and K^+) with water molecules (Pruteanu et al., 2016).

4.2. Similarity in classes of compounds

An own collection of compounds from plant species known to have anti-inflammatory potential has been considered and presented in Chapter 6. Similarity has been calculated between the natural compounds and a set of drug compounds used in the treatment of inflammatory disorders. Thus, the study wanted to highlight other natural sources of compounds with similar properties.

4.3. Quantitative structure-property relationships

To validate a quantitative structure-property relationship model was considered a set of 40 similar molecules (Pruteanu et al., 2016) from PubChem (<https://pubchem.ncbi.nlm.nih.gov/>) belonging to a class of steroids and their lipophilicity property expressed logarithmically by the octanol-water partition coefficient ($\log P$, usually <5) referring to the concentration ratio of the non-ionized species of a compound (Chapter 7).

4.4. Quantitative structure-property relationships

Molecular behavior was also observed at the cellular level, so following the treatment of three cell lines of arsenates breast cancer and a normal cell line (HUMEC) for control, changes in gene expression were investigated. It has been proposed to evaluate the effect of arsenates at the transcriptomic level and to investigate the mechanisms of action (activation of apoptosis, autophagy, reduction of cellular proliferation) to understand the behavior of the arsenates *in vitro* (Chapter 8).

4.5. Investigation of *Lactobacillus* therapeutic benefits *in vitro*

Regarding the therapeutic effects of organic compounds, a probiotic solution of three species of *Lactobacillus* was considered (Taranu et al., 2018). Following treatment of porcine intestinal epithelial cells (IPEC-1) with the mixture of lactobacilli was followed behavior of these compounds at gene expression level. (Chapter 9).

CHAPTER 5. HYDRATION MODELS OF MONOVALENT IONS

5.1. Materials and methods

In Table 5.1 are the distance values calculated by different methods (Experimental, MP4/6-31G*, MP3/6-31G*, MP2/6-31G*, M06-2X/6-31G*, HF/6-31G*, HF/3-21G*, HF/STO-3G) and the values closest to each other are given by the calculation: M06-2X/6-31G*, HF/3-21G* and MP3/6-31G* (Russell, 2006).

Table 5.1. The length d(O–H) in water: experimental vs. calculated (Russell, 2006)

Method	d(O–H) pm
Experimental	95.78
MP4/6-31G*	97.03
MP3/6-31G*	96.68
MP2/6-31G*	96.89
M06-2X/6-31G*	96.56
HF/6-31G*	94.73
HF/3-21G*	96.65
HF/STO-3G	98.92

5.1.2. Ion-water clusters

The workflow describing this procedure is detailed in the steps as follows:

- Step 1: water molecules have been placed in the vicinity of ions of interest
- Step 2*: geometry optimization was performed

➤ Step 3*: more water molecules have been added to the unoccupied spaces in the vicinity of the investigated ions

* Steps 2 and 3 were repeated until no change was observed in the arrangements of the groups of the first substrate.

The steps described above were applied for each ion under investigation.

5.1.3. Building congeners cages

The dodecahedron cages were constructed according to the methods described in the literature (Burnham et al., 2006; Grayson et al., 2009) for each investigated cation (NH_4^+ , Li^+ , Na^+ și K^+) and the structures obtained are presented in Figure 5.1.

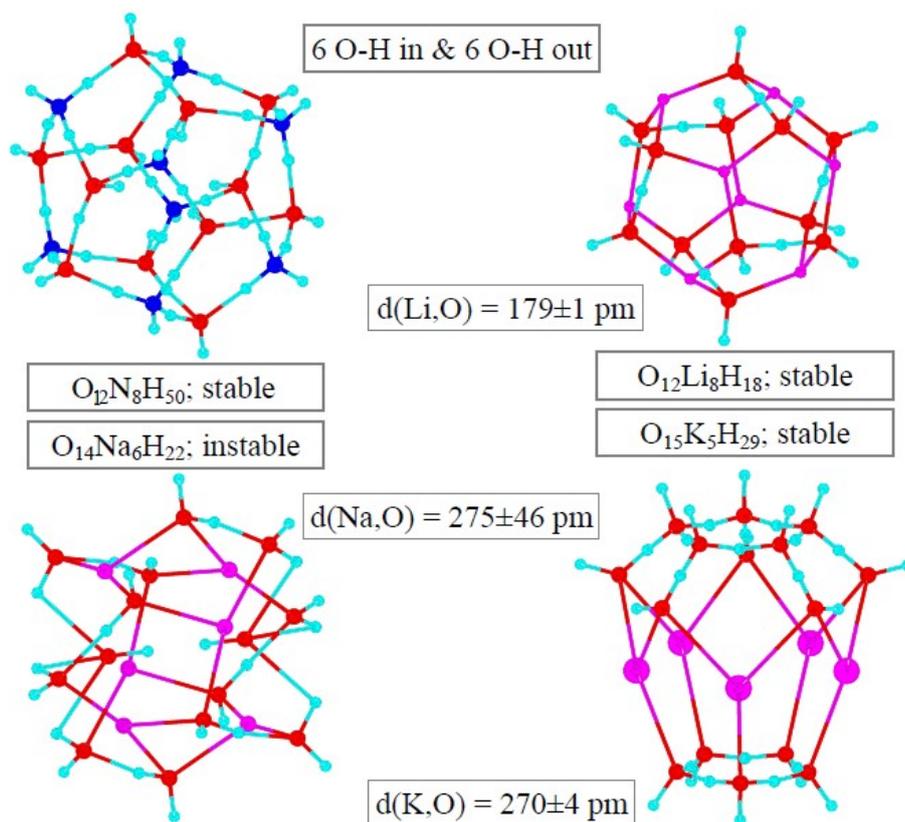


Figure 5.1. The dodecahedron groups for cations NH_4^+ , Li^+ , Na^+ and K^+ (O - red; H - blue; N - dark blue; Li, Na, K - pink)

Binding length averages were compared using the ANOVA statistical test with a level of significance 0.33% ($\alpha = 0.05$ adjusted by the number of comparisons considered, in this case the number being the number of ions investigated; $\alpha^* = 0.05/[6*(6-1)/2]$).

Considering Bonferroni test (Biella et al., 2008) the results showed significant differences. To test the differences between the angles formed between ion-water groups, the Friedman ANOVA test (Pruteanu et al., 2016), the statistical analysis being performed using the Statistics software (V.8.) (<http://software.dell.com/products/statistica/>).

5.1.4. Symmetry and molecular stability

In the case of ammonia ($O_{12}N_8H_{50}$) six O-H bonds are stabilized within the formed group and six O-H bonds are stabilized outside the structure (as can be seen in Figure 5.1).

The formation and / or stability of the dodecahedron groups integrating the water molecules were predominantly investigated when formed with ammonia.

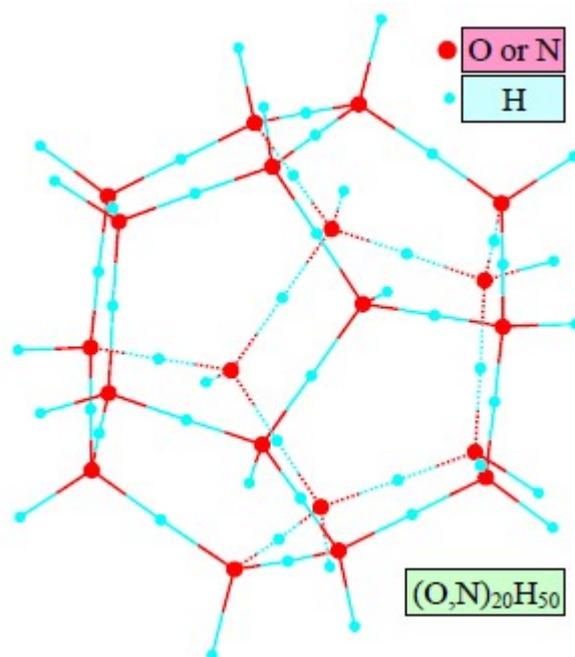


Figure 5.2. General structure of $O_xN_{20-x}H_{50}$ (O or N - red; H - blue)

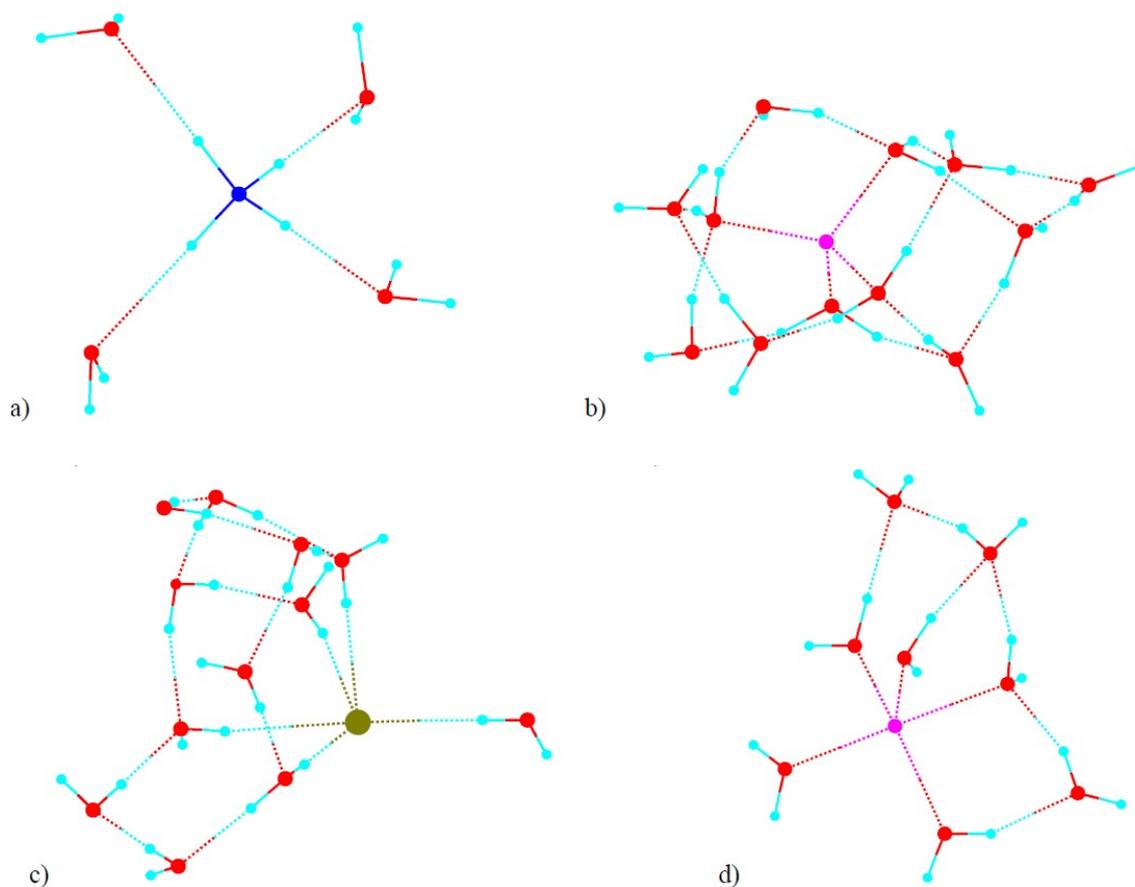
5.2. Results and discussions

5.2.1. Stability of ion-water clusters

The stable ion-water groups identified in the present case have been shown to be those formed with:

- four ($\text{NH}_4^+ \cdot 4\text{H}_2\text{O}$ and $\text{Li}^+ \cdot 4\text{H}_2\text{O}$);
- five ($\text{Cl}^- \cdot 5\text{H}_2\text{O}$ and $\text{Na}^+ \cdot 5\text{H}_2\text{O}$);
- six ($\text{F}^- \cdot 6\text{H}_2\text{O}$ and $\text{K}^+ \cdot 6\text{H}_2\text{O}$) water molecules.

The stable groups mentioned are represented in the figure below (Figure 5.3) with the mention that the ion-water groups are naturally formed and without constraints, since the silico model was carried out with the water molecules.



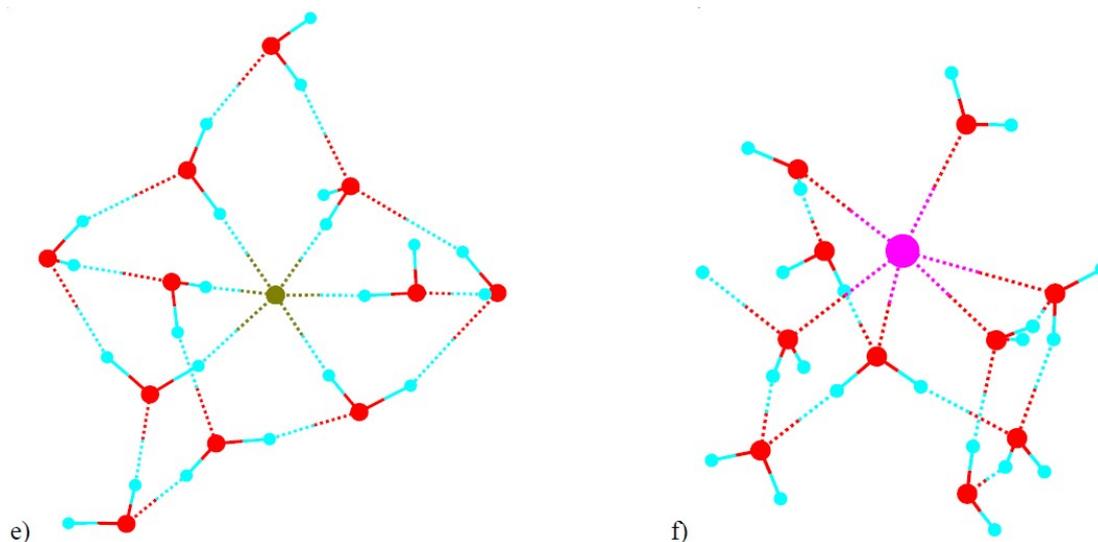


Figure 5.3. a) Group $\text{NH}_4^+ \cdot 4\text{H}_2\text{O}$, b) Group $\text{Li}^+ \cdot 4\text{H}_2\text{O}$, c) Group $\text{Cl}^- \cdot 5\text{H}_2\text{O}$, d) Group $\text{Na}^+ \cdot 5\text{H}_2\text{O}$, e) Group $\text{F}^- \cdot 6\text{H}_2\text{O}$, f) Group $\text{K}^+ \cdot 6\text{H}_2\text{O}$.

The first ion analyzed (NH_4^+) has been selected as a reference for the reproducibility of the calculation and for the validation of the analysis method, and the results obtained are in agreement with known data (Clegg și Brimblecombe, 1989; Galashev, 2013; Guerra et al., 2014; Janeiro-Barral și Mella, 2006).

The calculated bindings and angles appear to be very close to their values when naturally occurring in the aqueous environment. For $\text{Na}^+(\text{H}_2\text{O})_{20}$, $\text{Cl}^-(\text{H}_2\text{O})_{17}$, and $\text{Na}^+(\text{H}_2\text{O})_{100}$ using PBHaT algorithm (Burnham et al., 2006) has been identified at least globally. This algorithm is a hybrid capable of efficiently sampling the partition function from the global minimum to the liquid state.

5.2.2. Molecular symmetry

Analyzing the groups of water-ions, it is shown that the arrangement of the water molecule around the investigated ions is not symmetrical even if it appears to be symmetrical at first glance (see Figure 5.3a). Subsequent addition of water molecules, after the initial bond was created, it can be seen that asymmetry is increasingly evident.

It can be said that the molecular arrangement is expected to lose its symmetry when it comes to orbitals with higher energies. In fact, it is hard to believe that when water is dissolved in water in diluted solutions, the arrangement of water molecules will follow the expected symmetry of the fundamental ion state *in vitro*.

As expected, since different ion-water groups have been investigated, the lengths of the bonds and the angles between the bonds differ from one group to another. The results are shown in Tables (Table 5.2 and Table 5.3):

Table 5.2. Length of bonds and angles between bonds in the grouping $\text{Cl}^- \cdot 5\text{H}_2\text{O}$ and $\text{F}^- \cdot 6\text{H}_2\text{O}$

Distance		Angles			
d(Atom...Atom)	pm	(Atom...Atom) d(Atom...Atom)	(°)	(Atom...Atom) d(Atom...Atom)	(°)
Group $\text{Cl}^- \cdot 5\text{H}_2\text{O}$					
d(H...Cl)	265	(H...Cl) ₂₈₁ –(H...Cl) ₃₀₇	67	(H...Cl) ₂₇₄ –(H...Cl) ₂₈₁	96
d(H...Cl)	274	(H...Cl) ₂₇₄ –(H...Cl) ₃₀₇	76	(H...Cl) ₂₆₅ –(H...Cl) ₂₇₄	101
d(H...Cl)	281	(H...Cl) ₂₈₁ –(H...Cl) ₂₈₄	82	(H...Cl) ₂₆₅ –(H...Cl) ₂₈₄	129
d(H...Cl)	284	(H...Cl) ₂₇₄ –(H...Cl) ₂₈₄	82	(H...Cl) ₂₈₄ –(H...Cl) ₃₀₇	140
d(H...Cl)	307	(H...Cl) ₂₆₅ –(H...Cl) ₃₀₇	89	(H...Cl) ₂₆₅ –(H...Cl) ₂₈₁	146
Group $\text{F}^- \cdot 6\text{H}_2\text{O}$					
d(H...F)	179	(H...F) ₁₇₉ –(H...F) ₁₈₄	77	(H...F) ₁₈₄ –(H...F) ₁₈₅	89
d(H...F)	182	(H...F) ₁₈₄ –(H...F) ₁₈₇	85	(H...F) ₁₇₉ –(H...F) ₁₈₅	90
d(H...F)	183	(H...F) ₁₈₂ –(H...F) ₁₈₃	85	(H...F) ₁₈₃ –(H...F) ₁₈₇	93
d(H...F)	184	(H...F) ₁₇₉ –(H...F) ₁₈₃	86	(H...F) ₁₈₂ –(H...F) ₁₈₇	109
d(H...F)	185	(H...F) ₁₈₅ –(H...F) ₁₈₇	87	(H...F) ₁₇₉ –(H...F) ₁₈₇	162
d(H...F)	187	(H...F) ₁₈₃ –(H...F) ₁₈₄	88	(H...F) ₁₈₂ –(H...F) ₁₈₄	165
		(H...F) ₁₇₉ –(H...F) ₁₈₂	89	(H...F) ₁₈₃ –(H...F) ₁₈₅	177

In the case of anions Cl^- și F^- the bond between water molecules and ions is created with the contribution of water hydrogen ions, the region around the anions being rich in electrons. This explains why the distances and angles of the resulting molecular arrangements are given in relation to these hydrogen atoms (see Table 5.2 for Cl^- and F^-).

H...Cl⁻ distancs in grouping $\text{Cl}^- \cdot 5\text{H}_2\text{O}$ are close to those values of the distance between H...Cl⁻ in normal size clathrate, where high angles are possible (Laage și Hynes, 2007).

As can be seen for chloride ion, there is no symmetry due to the additional effect of the presence of free energy corresponding to d-type orbitals (see Table 5.2). Another finding is visible in the number of water molecules surrounding the anion.

5.2.3. Electronegativity

The difference in electronegativity between the fluoride ions and the chlorine ions is highlighted, the fluoride ion being able to attract six hydrogen atoms of the water molecules in the first layer, while the chlorine ion having lower electronegativity attracts only 5 hydrogen atoms water molecules.

By comparing the standard deviation between the angular values obtained and the expected values, it is noted that the standard deviation is twice as high as that obtained in the fluor-water ion group. For these calculations the platonic structure with five edges and six faces was considered ($6 \times 90^\circ$, $3 \times 120^\circ$, $1 \times 180^\circ$).

Taking into account the angles for the fluoride ion dissolved in water, the arrangement of the water molecules in the first layer surrounding the F^- ion is close to what is meant by a bi-pyramidal square arrangement (Table 5.2).

In the present case, because of the electronegativity difference between oxygen and fluoride ion, it can be seen that the arrangement is a bi-pyramidal square asymmetric type formed by the hydrogen atoms surrounding the fluoride ion.

For cations, things are exactly the opposite so that the bond between the water molecules and the cations is formed by the oxygen atoms. Due to this, and the situation where the number of coordination is decreasing with the increase of the atomic number observed in the case of the anions (6 for F^- , 5 for Cl^-), is reversed for cations (see Table 5.3).

Table 5.3. The length of the bonds and angles between the bonds in the cation-water groups: $NH_4^+ \cdot 4H_2O$, $Li^+ \cdot 4H_2O$, $Na^+ \cdot 5H_2O$, $K^+ \cdot 6H_2O$

Distance		Angles			
d(Atom...Atom)	pm	(Atom...Atom)	($^\circ$)	(Atom...Atom)	($^\circ$)
		d(Atom...Atom)		d(Atom...Atom)	
Group $NH_4^+ \cdot 4H_2O$					

Distance		Angles			
d(O...H)	95	(H...O) ₉₅ –(O...H) ₉₅	105		
d(N...H)	101	(H...N) ₁₀₁ –(N...H) ₁₀₁	109		
d(O...H)	208	(H...Cl) ₉₅ –(O...H) ₂₀₈	113		
Group Li⁺·4H₂O					
d(Li...O)	193	(Li...O) ₁₉₃ –(Li...O) ₁₉₆	99	(Li...O) ₁₉₃ –(Li...O) ₂₀₀	111
d(Li...O)	194	(Li...O) ₁₉₄ –(Li...O) ₂₀₀	104	(Li...O) ₁₉₄ –(Li...O) ₁₉₆	116
d(Li...O)	196	(Li...O) ₁₉₆ –(Li...O) ₂₀₀	110	(Li...O) ₁₉₃ –(Li...O) ₁₉₄	117
d(Li...O)	200				
Group Na⁺·5H₂O					
d(Na...O)	231	(Na...O) ₂₃₄ –(Na...O) ₂₃₉	84	(Na...O) ₂₃₄ –(Na...O) ₂₃₅	97
d(Na...O)	234	(Na...O) ₂₃₆ –(Na...O) ₂₃₉	85	(Na...O) ₂₃₁ –(Na...O) ₂₃₄	100
d(Na...O)	235	(Na...O) ₂₃₁ –(Na...O) ₂₃₉	88	(Na...O) ₂₃₄ –(Na...O) ₂₃₆	123
d(Na...O)	236	(Na...O) ₂₃₁ –(Na...O) ₂₃₅	93	(Na...O) ₂₃₁ –(Na...O) ₂₃₆	135
d(Na...O)	239	(Na...O) ₂₃₅ –(Na...O) ₂₃₆	94	(Na...O) ₂₃₁ –(Na...O) ₂₃₅	178
Group K⁺·6H₂O					
d(O...K)	278	(O...K) ₂₈₅ –(O...K) ₂₉₃	58.1	(O...K) ₂₇₈ –(O...K) ₂₈₄	94.4
d(O...K)	279	(O...K) ₂₈₁ –(O...K) ₂₈₅	79.0	(O...K) ₂₈₁ –(O...K) ₂₉₃	100.5
d(O...K)	281	(O...K) ₂₈₄ –(O...K) ₂₉₃	79.1	(O...K) ₂₇₉ –(O...K) ₂₈₄	121.6
d(O...K)	284	(O...K) ₂₇₈ –(O...K) ₂₈₁	80.8	(O...K) ₂₈₄ –(O...K) ₂₈₅	127.8
d(O...K)	285	(O...K) ₂₈₁ –(O...K) ₂₈₄	81.3	(O...K) ₂₇₈ –(O...K) ₂₈₅	128.8
d(O...K)	293	(O...K) ₂₇₉ –(O...K) ₂₉₃	86.2	(O...K) ₂₇₉ –(O...K) ₂₈₁	157.1
		(O...K) ₂₇₉ –(O...K) ₂₈₅	86.3	(O...K) ₂₇₈ –(O...K) ₂₉₃	173.0
		(O...K) ₂₇₈ –(O...K) ₂₇₉	92.5		

The average bond lengths between the investigated ion-water groups proved to be significantly different (ANOVA test with a p value, $p = 3 \cdot 10^{-10}$). The Bonferroni post-hoc test identified significant differences with respect to the binding lengths for the following pair of groups (differences were considered significant according to the significance level adjusted by 0.3333%):

- Average bond length for grouping $\text{NH}_4^+ \cdot 4\text{H}_2\text{O}$ proved to be less significant compared to that observed in the group $\text{Cl}^- \cdot 5\text{H}_2\text{O}$ (where $p=10^{-8}$), group $\text{Na}^+ \cdot 5\text{H}_2\text{O}$ ($p=8.7 \cdot 10^{-6}$), and group $\text{K}^+ \cdot 6\text{H}_2\text{O}$ ($p=5 \cdot 10^{-9}$).
- Average bond length for grouping $\text{Li}^+ \cdot 4\text{H}_2\text{O}$ proved to be less significant compared to that observed in the group $\text{Cl}^- \cdot 5\text{H}_2\text{O}$ ($p=2.3 \cdot 10^{-5}$), and group $\text{K}^+ \cdot 6\text{H}_2\text{O}$ ($1.6 \cdot 10^{-5}$).

- Average bond length for grouping $\text{Cl}^- \cdot 5\text{H}_2\text{O}$ proved to be significant compared to that observed in the group $\text{F}^- \cdot 6\text{H}_2\text{O}$ ($p=5 \cdot 10^{-7}$).
- Average bond length for grouping $\text{F}^- \cdot 6\text{H}_2\text{O}$ was found to be significantly lower compared to that observed in the group $\text{K}^+ \cdot 6\text{H}_2\text{O}$ ($p=2 \cdot 10^{-7}$).

The analysis of the angles presented in Table 5.2 and 5.3 led to the following:

- As expected, the smallest angle between the bonds was observed in a group of 6 water molecules, respectively, in the group $\text{K}^+ \cdot 6\text{H}_2\text{O}$ (angle=58.1°).
- No significant differences were observed when the angles between the links were statistically investigated (Statistica Friedman ANOVA=4.27 $p=0.5119$).

The results obtained in this case are associated with the concept of "infinite dilution", for which, for example, one can say that there are no other ions in the neighborhood (Table 5.4).

Table 5.4 contain:

- The ratio between water and ammonia for each group of this type (H_2O and NH_3);
- Number of hydronium ions (H_3O^+ column) and hydroxide ions (HO^- column) freed from group formation;
- Load rate released per total number of water molecules involved ($[\text{H}_3\text{O}^+]/\text{H}_2\text{O}$ column);
- The reaction leading to group formation (the "Training Reaction" column, which also represents the verification key for the previous calculations);
- The rate of nitrogen and oxygen atoms corresponding to the whole arrangement in the total mixture of water and ammonia.

Table 5.4. Conformations for $\text{O}_x\text{N}_{20-x}\text{H}_{50}$

No	Group	H_2O	NH_3	H_3O^+	HO^-	$[\text{H}_3\text{O}^+]/\text{H}_2\text{O}$	Reaction	$\text{N}/(\text{N}+\text{O})\%$
0	O_0N_{20}	0	20	10	0	$+(10)/(0+10)$	$20\text{NH}_3 + 10\text{H}_2\text{O} \rightarrow \text{O}_0\text{N}_{20}\text{H}_{50} + 10\text{H}_3\text{O}^+$	$18/28=100$
1	O_2N_{18}	2	18	8	0	$+(8)/(2+8)$	$18\text{NH}_3 + 10\text{H}_2\text{O} \rightarrow \text{O}_2\text{N}_{18}\text{H}_{50} + 8\text{H}_3\text{O}^+$	$18/28=64.3$

No	Group	H ₂ O	NH ₃	H ₃ O ⁺	HO ⁻	[[±]]/H ₂ O	Reaction	N/(N+O)%
2	O ₄ N ₁₆	4	16	6	0	+(6)/(6+4)	16NH ₃ + 10H ₂ O → O ₄ N ₁₆ H ₅₀ + 6H ₃ O ⁺	16/26=61.5
3	O ₅ N ₁₅	5	15	5	0	+(5)/(5+5)	15NH ₃ + 10H ₂ O → O ₅ N ₁₅ H ₅₀ + 5H ₃ O ⁺	15/25=60.0
4	O ₆ N ₁₄	6	14	4	0	+(4)/(4+6)	14NH ₃ + 10H ₂ O → O ₆ N ₁₄ H ₅₀ + 4H ₃ O ⁺	14/24=58.3
5	O ₈ N ₁₂	8	12	2	0	+(2)/(2+8)	12NH ₃ + 10H ₂ O → O ₈ N ₁₂ H ₅₀ + 2H ₃ O ⁺	12/22=54.5
6	O ₁₀ N ₁₀	10	10	0	0	+(0)/(0+10)	10NH ₃ + 10H ₂ O → O ₁₀ N ₁₀ H ₅₀	10/20=50.0
7	O ₁₂ N ₈	12	8	0	2	-(2)/(2+12)	8NH ₃ + 14H ₂ O → O ₁₂ N ₈ H ₅₀ + 2HO ⁻	8/22=36.4
8	O ₁₄ N ₆	14	6	0	4	-(4)/(4+14)	6NH ₃ + 18H ₂ O → O ₁₄ N ₆ H ₅₀ + 4HO ⁻	6/24=25.0
9	O ₁₅ N ₅	15	5	0	5	+(5)/(5+15)	5NH ₃ + 20H ₂ O → O ₁₅ N ₅ H ₅₀ + 5HO ⁻	5/25=20.0
10	O ₁₆ N ₄	16	4	0	6	+(6)/(6+16)	4NH ₃ + 22H ₂ O → O ₁₆ N ₄ H ₅₀ + 6HO ⁻	4/26=15.4
11	O ₁₈ N ₂	18	2	0	8	+(8)/(8+18)	2NH ₃ + 26H ₂ O → O ₁₈ N ₂ H ₅₀ + 8HO ⁻	2/28=07.1
12	O ₂₀ N ₀	20	0	0	10	+(10)/(10+20)	0NH ₃ + 30H ₂ O → O ₂₀ N ₀ H ₅₀ + 10HO ⁻	0/30=0.00

The angles between the ion-water groupings for the investigated ions were successfully obtained and the groups formed with a considerable number of water molecules can explain the dissolution of the ions investigated in water.

The "considerable number" of molecules or, in this case, referred to as the "magic number of clusters", to which they were investigated for:

- Li_nNa_{8-n}, Na_nK_{8-n}, and K_nLi_{8-n} (Fournier, 2008);
- (C₅H₅N)_n (H₂O)_m (n=1~2, m=1~4) (DeBlase et al., 2015);
- Methyl tert-butyl ether (MTBE) - water groups (Di Palma și Bende, 2013);
- H⁺(NH₃)(piridin)(H₂O)_n, H⁺(NH₃)(piridin)₂(H₂O)_n (n = 18, 20 și 27) (Ryding et al., 2012);
- H⁺(NH₃)₅(H₂O)₂₀ (tetrahedral ammonia encapsulated in a dodecahedron structure (H₂O)₂₀, found in clathrate) (Hvelplund et al., 2010).

For $\text{Na}^+(\text{H}_2\text{O})_{20}$, $\text{Cl}^-(\text{H}_2\text{O})_{17}$, and $\text{Na}^+(\text{H}_2\text{O})_{100}$ using PBHaT algorithm (Burnham et al., 2006) a global minimum was identified. This algorithm is a hybrid capable of efficiently sampling the partition function from the global minimum to the liquid state.

5.4. Conclusions

- Significant difference among congeners in the tendency of equilibrium arrangements as a result of interactions between water molecules (in this case the SM8 model was used); This algorithm is a hybrid capable of efficiently sampling the partition function from the global minimum to the liquid state.
- In the absence of other ions (at an infinite dilution) the arrangement is generally altered symmetry;
- In the presence of other ions (in concentrated solutions), the dodecahedron groups containing 8 lithium and 4 potassium atoms are symmetrical and stable while the dodecahedron groups containing 6 sodium atoms are unstable while the symmetry is altered.

CHAPTER 6. MOLECULAR SIMILARITY IN CLASSES OF ANTI-INFLAMMATORIES

6.1. Materials and methods

- Data collection of natural compounds with anti-inflammatory activity, from the literature;
Apply similarity measures with DataWarrior.

6.1.1. Data collection

Table 6.1. Information of interest collected for *Dipteracanthus prostratus* species

NCBI_ID Taxonomy	Compounds	SMILES	Uniprot_ID – Targets
1052855	Protocatechuic acid	<chem>OC(=O)c1ccc(c(c1)O)O</chem>	O75496, P22748, P23280,
	Gallic acid	<chem>OC(=O)c1cc(O)c(c(c1)O)O</chem>	P10145, P22748, P23280,

	Gallic acid methyl ester	COC(=O)c1cc(O)c(c(c1)O)O	P83916, P55789, B2RXH2,
--	--------------------------	--------------------------	-------------------------------

For example, experimental data show that 15 plant species belonging to the *Acanthaceae* family have been studied for their anti-inflammatory effect, and 39 genes belonging to the *Acanthaceae* family are known (Table 6.2).

Table 6.2. Plant species (36) belonging to the 8 families considered

Family	Gender/species	Taxonomy_NCBI_ID
Acanthaceae 42%	<i>Acanthus ilicifolius</i>	328098
	<i>Adhatoda vasica</i>	141317
	<i>Andrographis paniculata</i>	175694
	<i>Asteracantha longifolia</i>	883475
	<i>Asystasia gangetica</i>	141292
	<i>Barleria prionitis</i>	4189
	<i>Barleria lupulina</i>	101743
	<i>Dipteracanthus prostratus</i>	1052855
	<i>Elytraria acaulis</i>	640489
	<i>Nelsonia campestris</i>	4193
	<i>Phlogocanthus thyrsoiflorus</i>	526790
	<i>Pseuderanthemum palatiferum</i>	1685563
	<i>Rhinacanthus nasutus</i>	537489
	<i>Ruellia tuberosa</i>	441035
	<i>Thunbergia laurifolia</i>	504053
Achariaceae 8%	<i>Carpotroche brasiliensis</i>	1633205
	<i>Flacourtia indica</i>	210376
	<i>Gynocardia odorata</i>	124848
Agaricaceae 5%	<i>Agaricus blazei</i>	79798
	<i>Agaricus bisporus</i>	5341
Aizoaceae 6%	<i>Glinus oppositifolius</i>	764175
	<i>Trianthema portulacastrum</i>	3548
Alariaceae 3%	<i>Undaria pinnatifida</i>	74381
Alismataceae 3%	<i>Alisma plantago-aquatica</i> subsp. <i>Oriente</i>	262913
Altingiaceae 3%	<i>Liquidambar styraciflua</i>	4400
Amaranthaceae 30%	<i>Aerva javanica</i>	240009
	<i>Alternanthera philoxeroides</i>	381410
	<i>Alternanthera sessilis</i>	221762
	<i>Amaranthus spinosus</i>	124765

Family	Gender/species	Taxonomy_NCBI_ID
	<i>Celosia cristata</i>	124768
	<i>Salicornia herbacea</i>	259302
	<i>Suaeda maritima</i>	126913
	<i>Cyathula prostrata</i>	221766
	<i>Pupalia lappacea</i>	240105
	<i>Spinacia oleracea</i>	3562

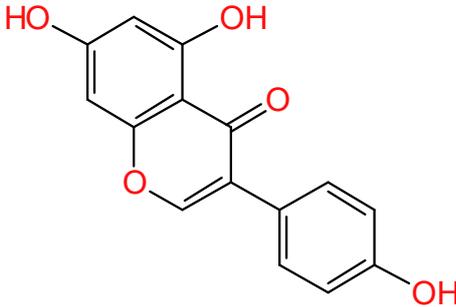
Anti-inflamatoarele ChEBI (în număr de 225 de compuși) au fost comparate cu un număr de 283 de compuși naturali extrași anterior din baza de date colectată (Table 6.4 din secțiunea Results and discussions) din 35 specii de plante ce aparțin a 8 familii (Table 6.2).

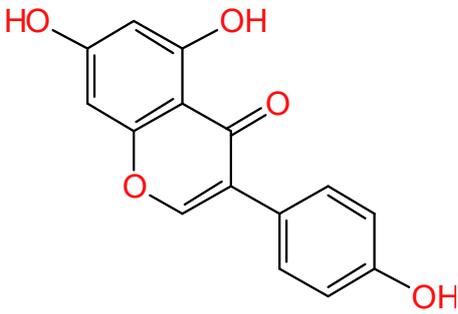
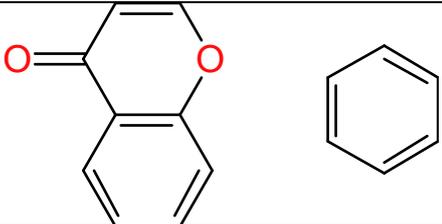
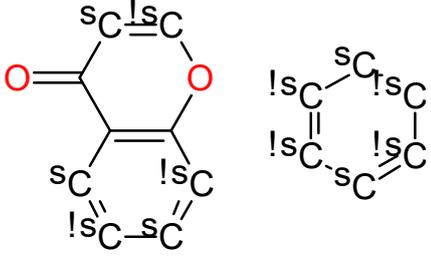
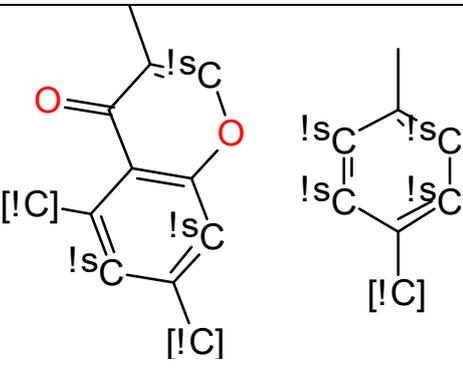
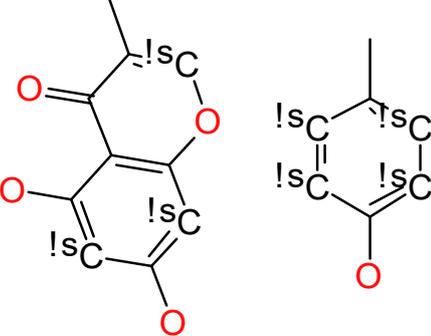
6.1.2. Scaffold based similarity

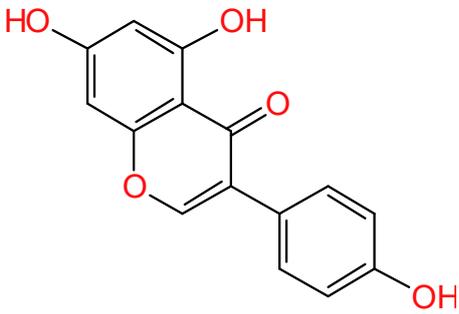
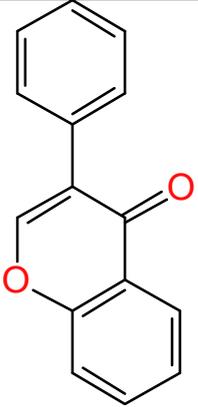
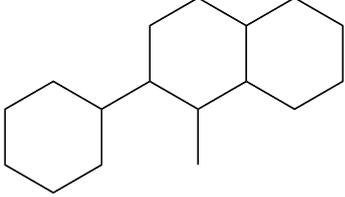
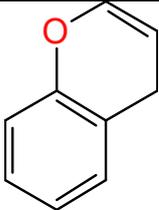
From the set of molecules, it is possible to highlight the central fragment common to most of the molecules, so substitutes are those that differ and confer different structural conformations.

For example, in Table 6.3, for the genistein molecule (DataWarrior representation), we can observe the types of molecular representations after which the calculations were performed to characterize the set of natural compounds collected.

Table 6.3. Molecular scaffolds representations taking as an example genistein molecule from the collection of natural compounds - representation from DataWarrior (Sander et al., 2015)

Genistein	
	
Scaffold	Representation

<p>Genistein</p> 	
<p>(a) Simple ring systems</p>	
<p>(b) Ring systems with substitution models s - substituents !s - exo-cyclic substituent, apart from hydrogen</p>	
<p>(c) Ring systems with substitution models carbon/hetero !s - exo-cyclic substituent, apart from hydrogen [!C] - any atom substituent apart from H and C</p>	
<p>(d) Ring systems without atomic substituents !s - exo-cyclic substituent, apart from hydrogen</p>	

Genistein	
(e) Scaffold/Murcko skeleton	
(f) Murcko skeleton graph	
(g) Central ring system	

With DataWarrior, similarity has been calculated in different ways depending on the purpose and the method we wanted to apply. Starting from the simplified computational method of molecular similarity in calculating similarity in 3D space that takes into account both the geometry and the way of binding the atoms.

The result was comparing molecules through these descriptors specific to each molecule in the set. The resulting results (similarity score) based on Euclidean SOM calculations (Self Organizing Map) (Sander et al., 2015) and the Tanimoto calculation

(Willett, 2011), show how many common compounds have in common, and how similar the two molecules are (section Results and discussions).

6.1.3. Principal component analysis and data screening

Based on scaffolds, similarity graphics such as PCA (Principal component Analysis) were built into DataWarrior. Also, the molecules in the vicinity of a major molecule were generated from similarity point of view.

To calculate the similarity, the Tanimoto coefficient was considered, and by the descriptors considered, according to (Sander et al., 2015):

- Binary fragments of substructure fragments (FragFp) were used;
- Stereochemistry was considered, duplicate fragments were counted and heteroatoms encoded (SkelSpheres).

The work algorithm was considered in the following steps:

- Positioning the set of molecules in 2D space;
- Calculating the matrix of similarity between all molecules;
- Locating the most similar neighboring molecules;
- Between two neighboring molecules, highlighting the attraction forces that grow with the similarity score and the distance between them;
- Visualization of similarity and similar molecules.

6.2. Results and discussions

6.2.2. Similarity clusters

The visualization of similarity relationships is highlighted based on clusters formed by correlation with more similar neighborhood components. The data was generated using DataWarrior software following similarity calculations comparing the two sets of annotated molecules ANTIINFLAMMATORY (blue) and CHEBI (red) (Figure 6.2).

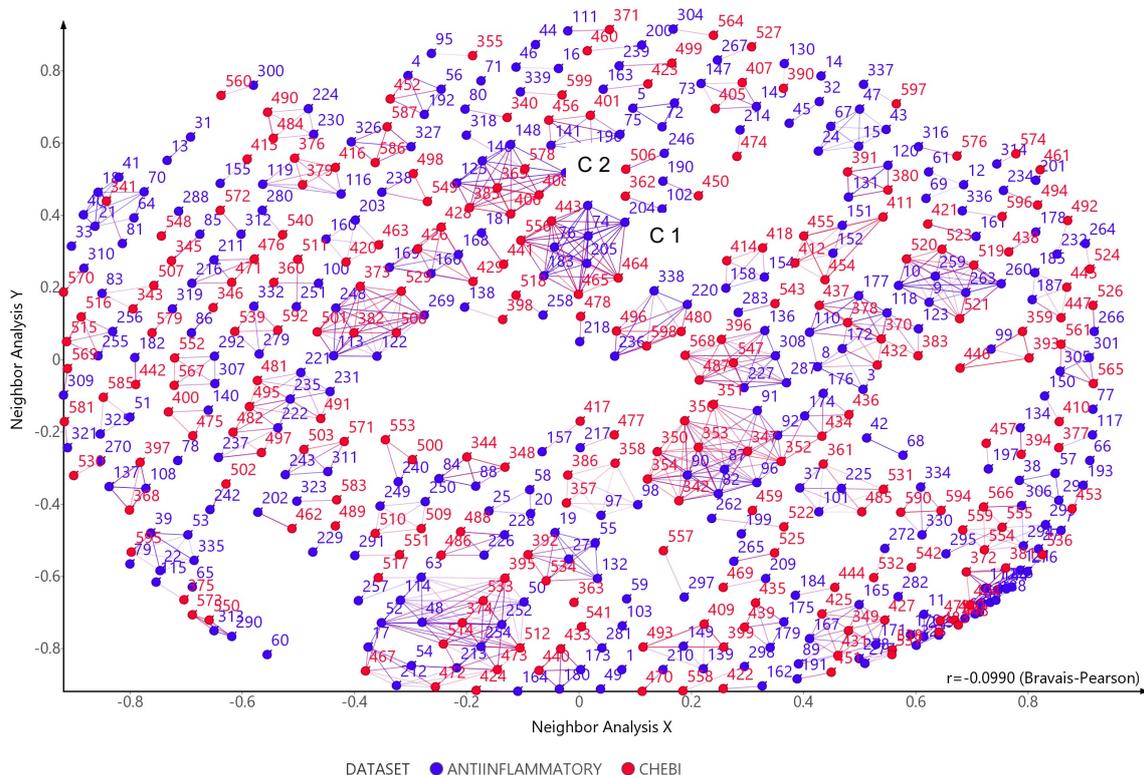


Figure 6.2. Similarity clusters resulting from neighborhood similarity calculation (DataWarrior view) - cluster 1 (C1), cluster 2 (C2) example.).

This correlation of the similarity between the data sets was made on the basis of the calculation of the Bravais-Pearson coefficient, resulting in a correlation coefficient very close to the ideal value 1, $r = 0.0990$ (Figure 6.2).

Clusters formed as can be seen in Figure 6.2 (example cluster 1 composed of molecules corresponding to the IDs 443, 556, 441, 74, 205, 76, 183, 518, 258, 465, 464, such as cluster 2 composed of molecules corresponding IDs: 146, 125, 549, 428, 38, 36, 578, 408, 400, 181) are subgroups of molecules connected to each other based on the neighborhood belonging to both analyzed datasets, indicating a good correlation between them in terms of similarity.

Based on structural descriptors (SMILES) and SkelSpheres, while also considering the structural scaffolds of the type of simple cyclic systems corresponding to the components, similarity was observed in the clusters formed (Figure 6.3 and Figure 6.4).

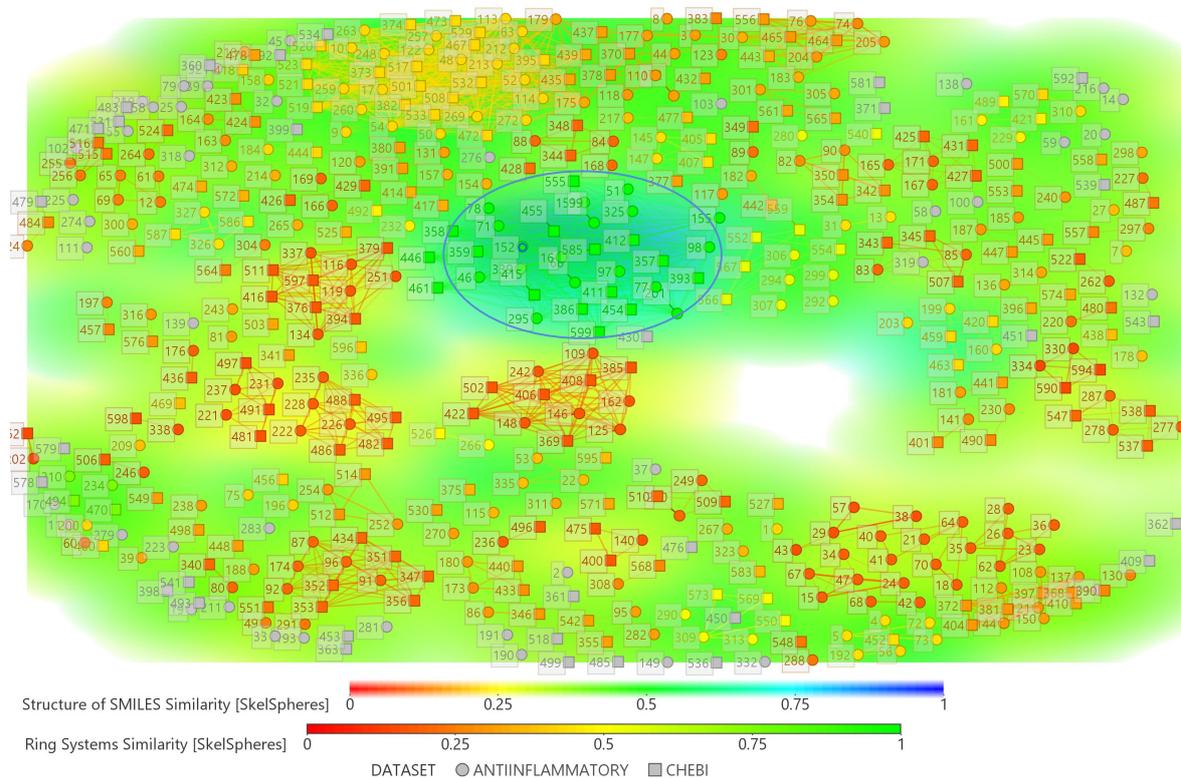


Figure 6.3. Similarity based on Ring Systems, SMILES and SkelSpheres descriptors, forms clusters whose components are very similar in structural terms. Similarity to the main leader (leader) ID 152 (see Supplementary Table 6.1) of the set of natural compounds.

A high degree of similarity is highlighted having as a leader (main components) of comparison, one molecule in each cluster. In Figure 6.3, the main component is the ID number 152 corresponding to vanillic acid in the set of natural compounds (ANTIINFLAMMATORY).

In Figure 6.4, the main component is the ID number 109 corresponding to the oleanolic acid in the set of natural compounds. The similarity relationships of the components of the two mentioned clusters are shown in Table 6.5.

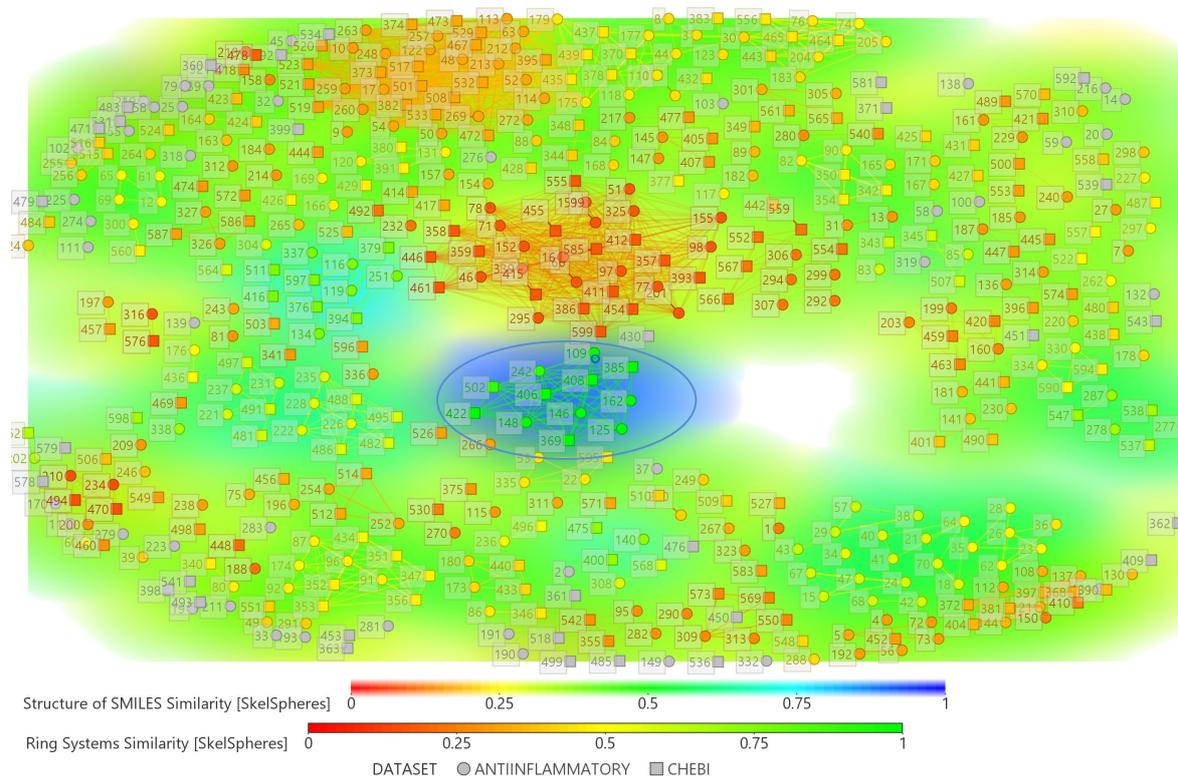
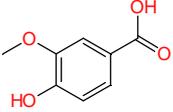
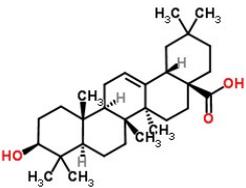
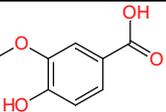
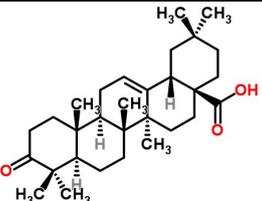
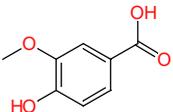
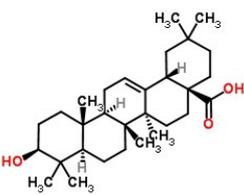
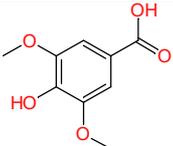
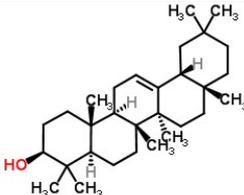
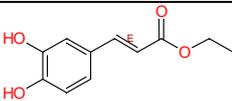
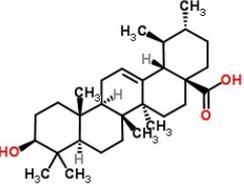
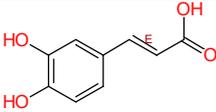
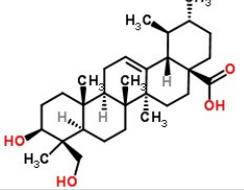
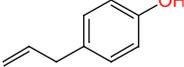
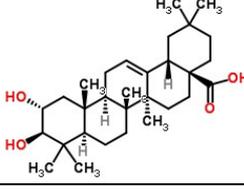
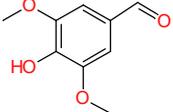
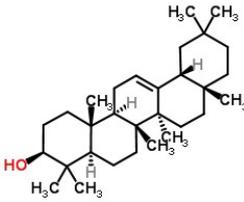
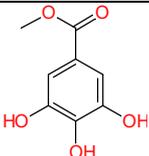
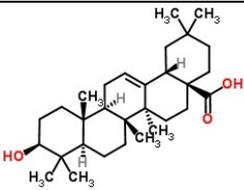
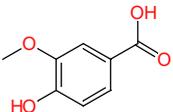
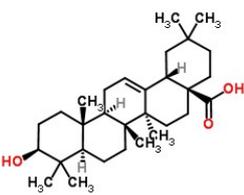
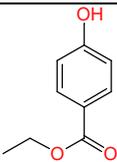
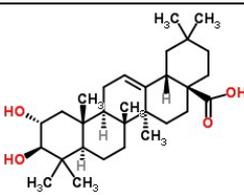
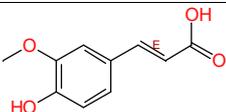
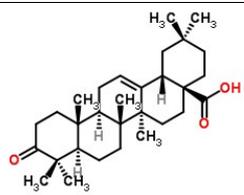
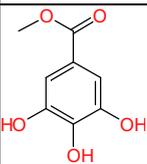
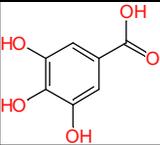
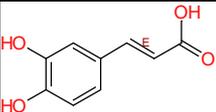
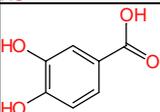
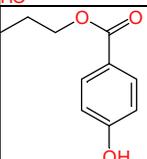


Figure 6.4. Similarity based on Ring Systems, SMILES and SkelSpheres descriptors, forms clusters whose components are very similar in structural terms. Similarity to the main (leader) ID 109 (see Supplementary Table 6.1) of the set of natural compounds.

Table 6.5. The similarity relationships of the components of the two similarity clusters (with the appropriate leaders considered) representatives of Figures 6.3 and 6.4.

Cluster leader 1 (ANTIINFLAMMATORY:A)				Cluster leader 2 (ANTIINFLAMMATORY:A)			
152				109			
ID	Structure 2D	Score	Set	ID	Structure 2D	Score	Set
455		1	CHEBI	242		1	A

Cluster leader 1 (ANTIINFLAMMATORY:A)				Cluster leader 2 (ANTIINFLAMMATORY:A)			
							
152				109			
454		1	CHEBI	162		1	A
599		1	CHEBI	406		0.934	CHEBI
585		1	CHEBI	385		0.933	CHEBI
555		1	CHEBI	408		0.927	CHEBI
461		1	CHEBI	422		0.910	CHEBI
446		1	CHEBI	369		0.900	CHEBI

Cluster leader 1 (ANTIINFLAMMATORY:A)				Cluster leader 2 (ANTIINFLAMMATORY:A)			
							
152				109			
46		0.916	A	148		0.877	A
66		0.887	A	502		0.877	CHEBI
99		0.858	A				
98		0.849	A				
325		0.845	A				
155		0.839	A				
71		0.829	A				

Taking into account one leader component of each distinct similarity cluster, it is noticed how global similarity calculated considering neighborhood changes in agreement with them (152 and 109). By focusing on the two clusters with a similarity score almost equal to 1, we can see the similarity between the compounds of the two

sets (CHEBI and natural anti-inflammatory compounds). Thus the most similar compounds from the two clusters can be found in Table 6.5.

6.2.3. Self organizing map

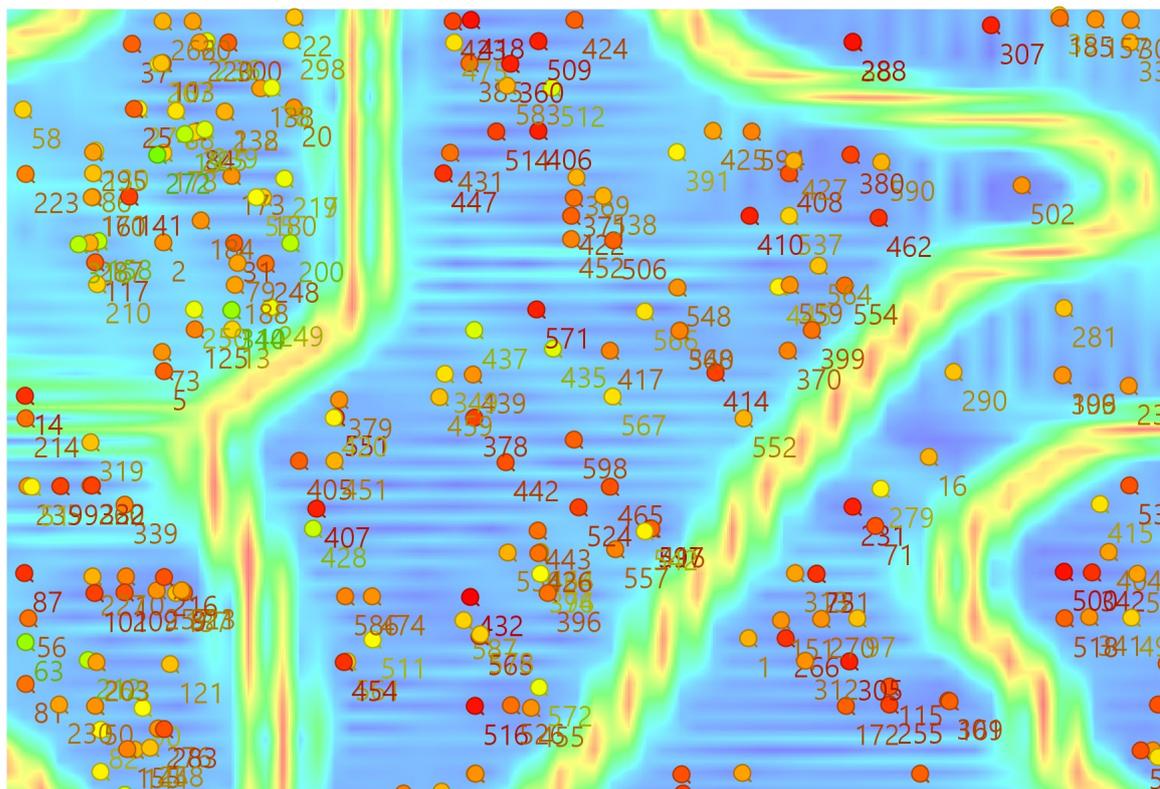


Figure 6.5. Self Organizing Map for arranging molecules according to the similarity of scaffolds to the compounds and viewing similarity clusters on the basis of the SkelSpheres neighborhood and descriptors (see DataWarrior).

By SOM we have taken into account the exact similarity values (Figure 6.5), and in the case of PCA, the scaffolds are separated according to two or three main components, as can be seen in the following (Fig. 6.6, Figure 6.7, Figure 6.8).

6.2.4. Principal component analysis

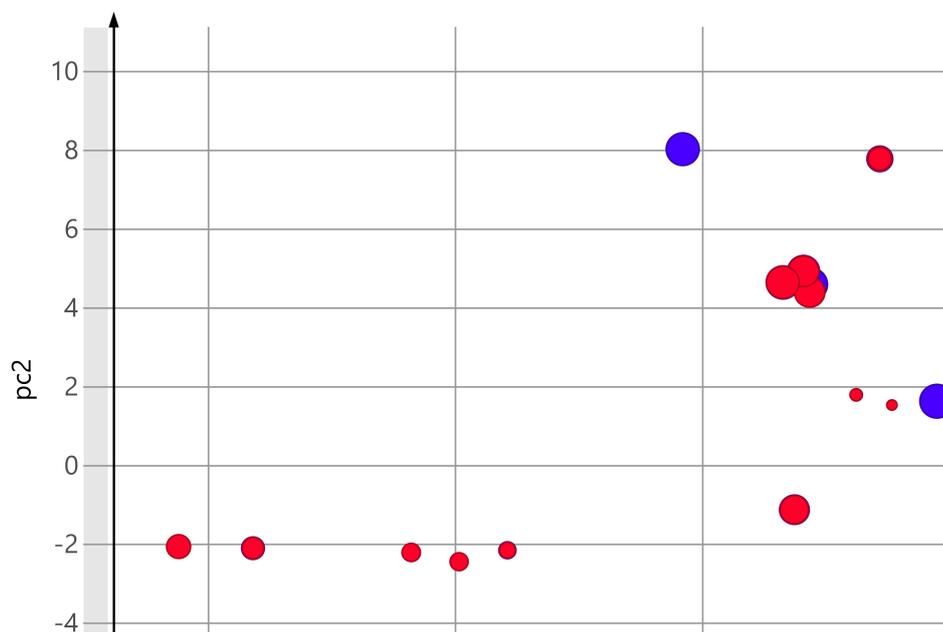


Figure 6.6. Analysis of the main components in the vicinity of Y, following the similarity calculation using the molecular scaffold and FragFp (according to the DataWarrior result) with a 23.2% variation of the main component 1 and 24.6% of the main component 2.

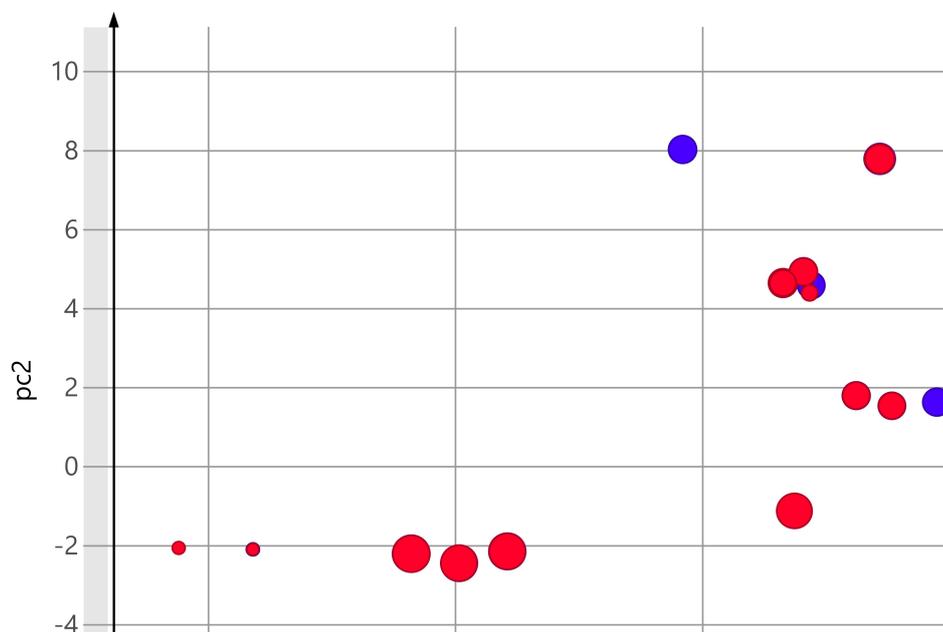


Figure 6.7. Analysis of the main components in the vicinity of X, following the similarity calculation using the molecular scaffold and FragFp (according to the DataWarrior result), with a 35.6% variation of the main component 1 and 40.2% of the main component 2.

Following the two figures (Figure 6.6 and Figure 6.7), one can say that there is a significant similarity between the molecules and that there is a small number of molecules that are not similar to them. Considering stereochemistry, the SkelSpheres descriptor counted the duplicate fragments and encoded the heteroatoms by calculating similarity based on this descriptor (Figure 6.8).

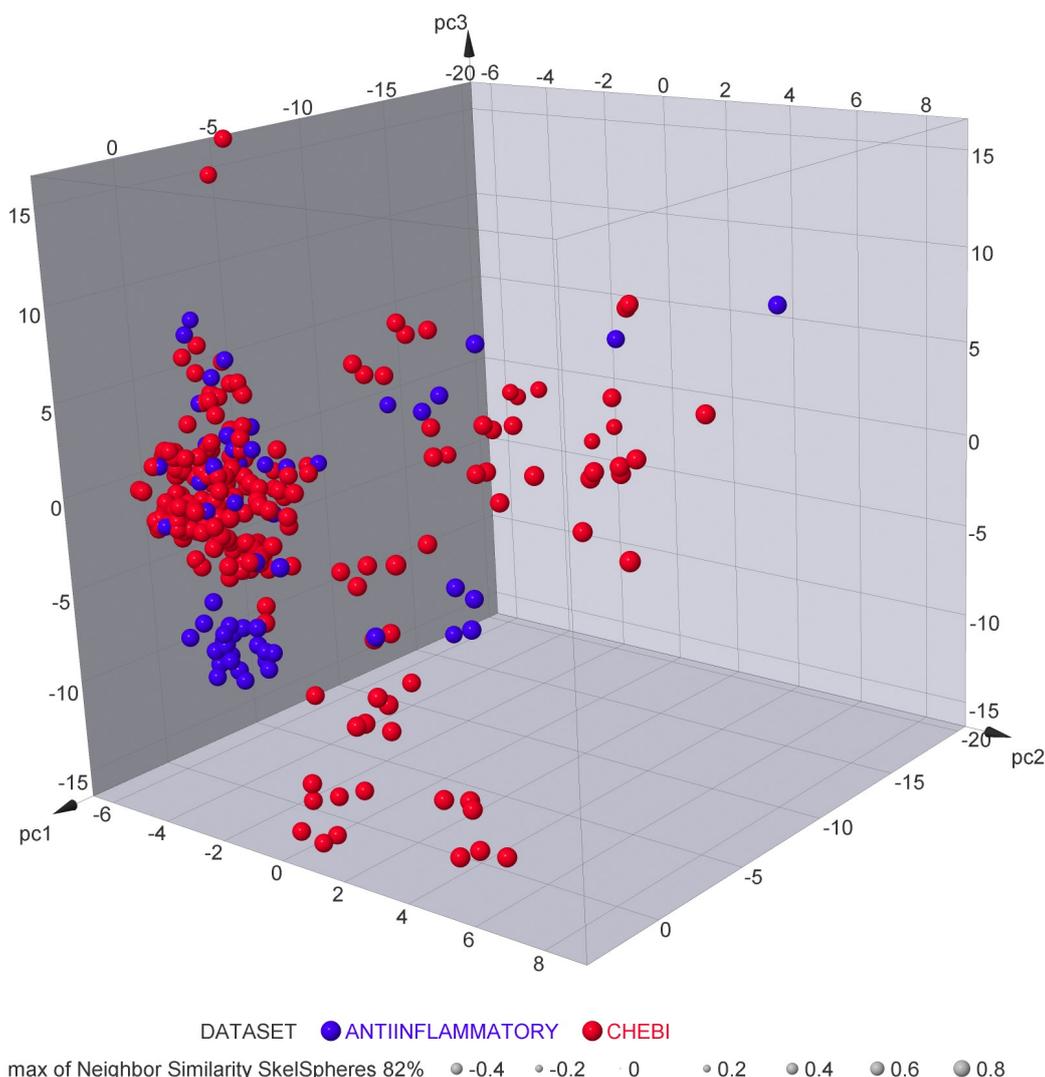


Figure 6.8. 3D view of the similarity of the molecules of the two sets of data (ANTIINFLAMMATORY - blue, CHEBI - red) and its assessment through the main components analysis (CP) using the SkelSpheres descriptor in the 3D space.

In the main component analysis (PCA) in 3D space, the similarity between components is confirmed by the graph in Figure 6.8. Similarity is thus rendered by the

statistical significance given by the result of the similarity score of 82% with values between 0.6 and 0.8.

6.4. Conclusions

- A series of natural compounds in the collected data set are similar to the structures of the drug compounds in the set taken as a reference;
- A chemical compound and its derivative can be found in a variety of plant extracts;
- The results contribute to the periodic confirmation and recurrence of compounds found to have anti-inflammatory activity.

CHAPTER 7. QUANTITATIVE STRUCTURE-PROPERTY RELATIONSHIP ON STEROIDS

7.1. Materials and methods

Table 7.1. The 40 derivatives of the 7 β -hydroxysteroid molecule

No.	PubChem ID	logP									
1	12760132	10.2	11	76310266	8.2	21	57390981	5.2	31	76322257	10.7
2	70682679	7.1	12	56663807	6.4	22	16758147	8.5	32	76325907	3.8
3	70682680	6.5	13	56847117	6.2	23	22213946	6.2	33	76327928	3.8
4	70688976	6.2	14	70686910	6	24	16759984	5.9	34	76333144	4.2
5	70693211	6.2	15	70691082	7.1	25	16758161	4.2	35	371617	6.1
6	70697302	6.5	16	11647965	8.4	26	76336739	11.2	36	313039	8
7	12836861	4.7	17	52947587	4.9	27	57396177	3	37	9922115	4.2
8	24867469	4.2	18	24982302	3.8	28	57399636	3	38	9924252	5.4
9	16082386	8.1	19	49823443	5.9	29	57401396	3	39	11551321	3
10	12358742	5.2	20	22216291	3	30	76322252	9.8	40	11957457	4.2

7.1.1. Geometry optimization

Molecular geometries have been optimized in the HyperChem program at semi-empirical PM3 (Parameterized Model No. 3). The log data files with data collection were extracted using the JSChem utility (Harsa et al., 2014).

7.1.2. Building the hypermolecule

Based on each position of the atoms that form the hyper molecule (see Figure 7.1), the binary vectors and corresponding mass fragments for each molecule in the set were calculated.

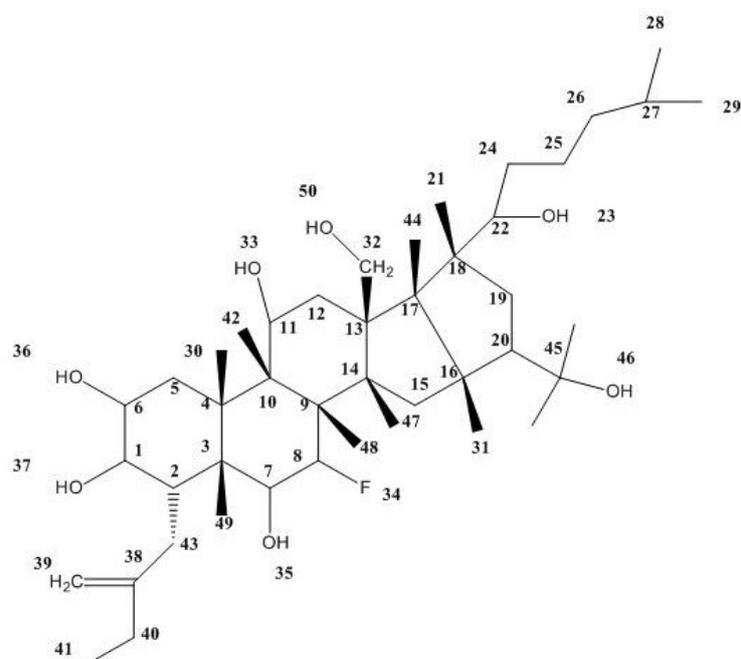


Figure 7.1. Hypermolecule formed by the overlap of forty ligands. 2D Graphic representation and graph numbering (Pruteanu et al., 2016) in ChemBioDraw („ChemBioDraw 14.0 User Guide”, 2016).

7.1.3. Topological descriptors

Table 7.2. Topology indices calculated for 7 β -hydroxysteroid derivatives in Table 7.1

Mol.	AD	CON	Di	D3D	De	CjDi	CjDe	CFDi	CFDe
1	35	35	2522	4138.48	7313	4573.5	1750.5	5062	1809
2	34	34	2670	3980.77	6965	4467	1923.5	4918	1966.5
3	33	33	2369	3510.96	6451	4047	1665	4487	1706.5
4	33	33	2342	3522.35	6424	4019.5	1638	4460	1679.5
5	33	33	2342	3355.66	6424	4019.5	1638	4460	1679.5
6	33	33	2369	3520.97	6451	4047	1665	4487	1706.5
7	25	25	926	1354.38	3297	1729.5	546.5	1983	572.5
8	24	24	802	1163.2	2969	1504.5	463	1739	486.5
9	33	33	2335	3350.73	6511	4029	1648.5	4550.5	1693
10	26	26	1052	1868.14	3627	1956.5	632	2229	660.5

Mol.	AD	CON	Di	D3D	De	CjDi	CjDe	CFDi	CFDe
11	35	35	2237	3659.41	8399	4455.5	1302	5281	1361
12	33	33	2345	3436.73	6348	3996	1647.5	4362	1687.5
13	33	33	2342	3309.13	6424	4019.5	1638	4460	1679.5
14	32	32	2098	3117.36	5967	3657.5	1436.5	4086	1476.5
15	34	34	2560	3759.51	6855	4349.5	1813.5	4808	1856.5
16	36	36	2435	3417.31	8962	4849.5	1424.5	5718	1485
17	24	24	796	1153.47	2949	1490.5	461	1732	488
18	27	27	1149	1681.56	3971	2130	715.5	2485	747.5
19	23	23	699	988.57	2662	1310	402	1516	421.5
20	25	25	887	1279.08	3279	1655	535.5	1966	562.5
21	26	26	1052	1530.66	3627	1956.5	632	2229	660.5
22	36	36	2436	3497	9036	4811.5	1459.5	5703.5	1521.5
23	33	33	2342	3355.66	6424	4019.5	1638	4460	1679.5
24	23	23	699	1023	2662	1310	402	1516	421.5
25	24	24	802	1158.41	2969	1504.5	463	1739	486.5
26	35	35	2944	4219.49	7659	5068.5	2080.5	5527	2130
27	25	25	887	1279.08	3279	1655	535.5	1966	562.5
28	25	25	887	1279.08	3279	1655	535.5	1966	562.5
29	25	25	887	1279.08	3279	1655	535.5	1966	562.5
30	34	36	2668	3785.72	7103	4609	1864.5	5048.5	1911.5
31	34	34	2668	3828.26	7103	4609	1864.5	5048.5	1911.5
32	27	27	1172	1646	3967	2178	728	2483	763
33	27	27	1172	1654.23	3967	2177.5	728	2482.5	762.5
34	24	24	802	1169.48	2969	1504.5	463	1739	486.5
35	24	24	732	1086.05	2660	1437	483.5	1707.5	501.5
36	32	32	2276	3204.88	6075	3832	1619	4173	1656.5
37	24	24	802	1169.48	2969	1504.5	463	1739	486.5
38	28	30	1412	2008.12	4395	2535	914	2829	944.5
39	25	25	887	1279.08	3279	1655	535.5	1966	562.5
40	24	24	802	1169.48	2969	1504.5	463	1739	486.5

7.1.4. Significant atom positions

Table 7.3. Statistically significant positions correlated with the mass of the fragments

Mol.	p17	p18	p26	p28	p33	p34	p35	p36	p37	p40	p43	p50
1	12.011	12.011	12.011	0	0	0	0	0	17.007	0	12.011	0
2	12.011	12.011	12.011	12.011	0	0	17.007	0	17.007	0	0	0
3	12.011	12.011	12.011	12.011	0	0	17.007	0	17.007	0	0	0
4	12.011	12.011	12.011	0	0	0	17.007	0	17.007	0	0	0
5	12.011	12.011	12.011	0	0	0	17.007	0	17.007	0	0	0
6	12.011	12.011	12.011	12.011	0	0	17.007	0	17.007	0	0	0

Mol.	p17	p18	p26	p28	p33	p34	p35	p36	p37	p40	p43	p50
7	12.011	17.007	0	0	0	0	0	0	17.007	0	0	0
8	12.011	17.007	0	0	0	0	0	0	17.007	0	0	0
9	12.011	12.011	12.011	0	0	19	17.007	0	17.007	0	0	0
10	12.011	12.011	0	0	0	0	0	0	17.007	0	0	0
11	12.011	12.011	0	0	0	0	0	0	17.007	0	12.011	0
12	12.011	12.011	12.011	0	0	0	0	0	17.007	0	0	0
13	12.011	12.011	12.011	0	0	0	17.007	0	17.007	0	0	0
14	12.011	12.011	12.011	0	0	0	17.007	0	17.007	0	0	0
15	12.011	12.011	12.011	12.011	0	0	17.007	0	17.007	0	0	0
16	12.011	12.011	0	0	0	0	0	0	17.007	0	12.011	0
17	12.011	0	0	0	0	0	0	17.007	17.007	0	0	0
18	12.011	12.011	0	0	17.007	0	0	0	17.007	0	0	0
19	12.011	0	0	0	0	0	0	0	17.007	0	0	0
20	12.011	17.007	0	0	17.007	0	0	0	17.007	0	0	0
21	12.011	12.011	0	0	0	0	0	0	17.007	0	0	0
22	12.011	12.011	0	0	0	0	17.007	0	0	0	0	0
23	12.011	12.011	12.011	0	0	0	17.007	0	17.007	0	0	0
24	12.011	0	0	0	0	0	0	0	17.007	0	0	0
25	12.011	17.007	0	0	0	0	0	0	17.007	0	0	0
26	12.011	12.011	12.011	0	0	0	0	0	17.007	12.011	12.011	0
27	12.011	17.007	0	0	17.007	0	0	0	17.007	0	0	0
28	12.011	17.007	12.011	0	17.007	0	0	0	17.007	0	0	0
29	12.011	17.007	0	0	17.007	0	0	0	17.007	0	0	0
30	12.011	12.011	12.011	0	0	0	0	0	17.007	0	12.011	0
31	12.011	12.011	12.011	0	0	0	0	0	17.007	12.011	12.011	0
32	12.011	12.011	0	0	0	0	0	0	17.007	0	0	17.007
33	12.011	12.011	0	0	0	0	0	0	17.007	0	0	17.007
34	12.011	17.007	0	0	0	0	0	0	17.007	0	0	0
35	0	0	0	0	0	0	17.007	17.007	17.007	0	0	0
36	12.011	12.011	12.011	17.007	0	0	0	0	17.007	0	0	0
37	12.011	17.007	0	0	0	0	0	0	17.007	0	0	0
38	12.011	12.011	0	0	0	0	0	0	17.007	0	0	0
39	12.011	17.007	0	0	17.007	0	0	0	17.007	0	0	0
40	12.011	17.007	0	0	0	0	0	0	17.007	0	0	0

The logP property was modeled using mass fragments as structural features of the molecules in the chosen set. The model has been validated through "leave-one-out" and "training vs. test" proceedings.

7.2. Results and discussions

7.2.1. Regression model based on significant atom positions

A significant regression model was obtained by which seven variables were identified as significant positions (Eq.1 is shown in Table 7.4).

$$\log P = 36.2431 + 0.0180 \cdot CjDi - 1.6780 \cdot AD - 0.0353 \cdot Di + 0.0228 \cdot CjDe - 0.0605 \cdot P18 - 0.0542 \cdot P33 - 0.0495 \cdot P35 \quad (\text{Eq. 1})$$

$R^2 = 0.9610$, $R^2_{adj} = 0.9525$, $Q^2 = 0.9413$; $s = 0.4808$, $n = 40$

F-Statistica (p-value) = 113 ($1.02 \cdot 10^{-20}$) where,

R^2 = determination coefficient,

R^2_{adj} = determination coefficient adjusted,

Q^2 = determination coefficient in leave-one-out proceeding,

s = estimated standard error,

n = sample size; F-Statistica = Statistica Fisher,

p-value = the probability of obtaining a significant model.

Table 7.4. Significant positions and their regression coefficients

Variables	Coefficients	Standard error	t Stat (p-value)
Interception	36.2431	4.4755	8.10 ($3.01 \cdot 10^{-9}$)
CjDi	0.0180	0.0020	8.82 ($4.41 \cdot 10^{-10}$)
AD	-1.6780	0.2383	-7.04 ($5.53 \cdot 10^{-8}$)
DI	-0.0353	0.0060	-5.85 ($1.69 \cdot 10^{-6}$)
CjDe	0.0228	0.0043	5.35 ($7.12 \cdot 10^{-6}$)
P18	-0.0605	0.0197	-3.07 ($4.37 \cdot 10^{-3}$)
P33	-0.0542	0.0157	-3.44 ($1.63 \cdot 10^{-3}$)
P35	-0.0495	0.0137	-3.61 ($1.04 \cdot 10^{-3}$)

CjDi = distance Cluj; AD = Adjacency;

Di = Distance; CjDe = Cluj detour;

P18 = Position 18; P33 = Position 33; P35 = Position 35

The model with the lowest number of predictors was chosen as the model with the most explanatory power. This was achieved by successive and repeated application of the step-by-step method for the set of descriptors in Tables 7.3 and 7.4.

7.2.2. Leave-one-out validation

For the validation of the model, a leave-one-out analysis was performed with a determination coefficient in the leave-one-out analysis $Q^2 = 0.9413$ (see Eq.1).

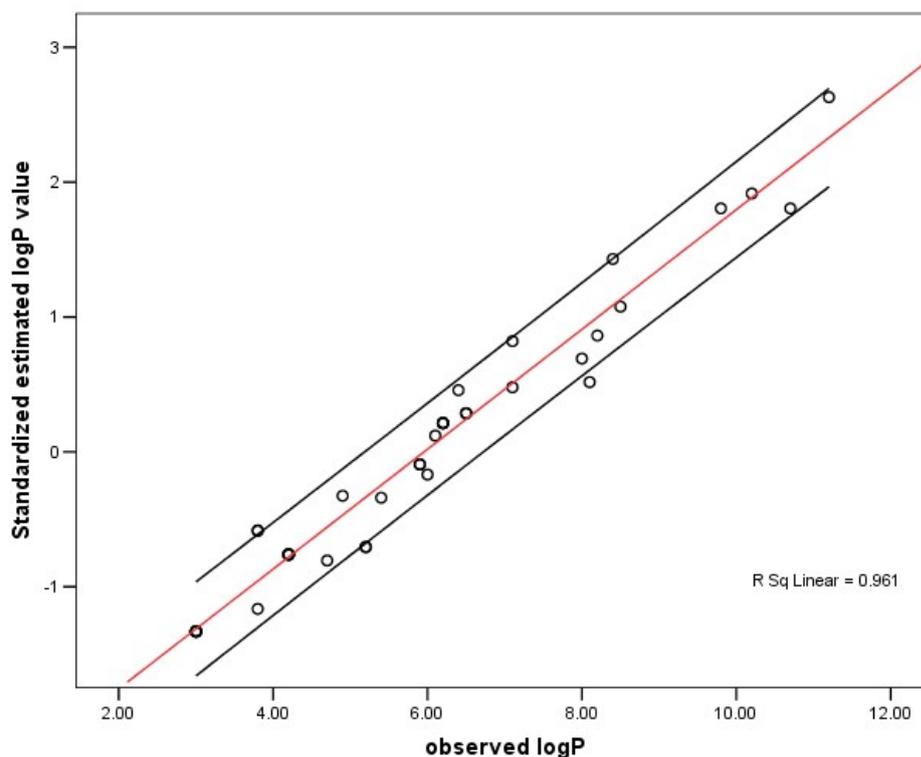


Figure 7.2. The estimated model (the red line - the pattern obtained and the black lines represent the confidence interval of 95%)

7.2.3. Training vs. test validation

The resulting model was also validated through the training vs. test on the set of 40 analyzed molecules (Table 7.1) for the set of descriptors.

The set of 40 molecules was divided into two sets, 24 molecules in the training set and the remaining 16 molecules in the test set.

The following molecules were randomly chosen to be part of the training set: 57396177, 49823443, 16082386, 16758147, 22216291, 9922115, 70688976, 70682680, 22213946, 11647965, 57390981, 12358742, 313039, 76325907, 57401396, 76327928, 76322252, 16759984, 24982302, 52947587, 12760132, 76336739, 76310266 și 70697302.

The regression equation obtained with the training molecule set was used to predict the logP values for the rest of the test set molecules:

$$\log P_{\text{train}} = 27.014 - 0.08755 \cdot p18 - 0.03625 \cdot p33 - 0.06206 \cdot p35 - 1.2697 \cdot AD - 0.001154 \cdot DI + 0.006573 \cdot CjDi \text{ (Eq.2)}$$

$$R^2_{\text{train}} = 0.9337$$

$$F_{\text{train}} = 40 \text{ (} p_F < 5 \cdot 10^{-9} \text{)}$$

$$R^2_{\text{test}} = 0.873$$

$$F_{\text{test}} = 9 \text{ (} p_F < 2.6 \cdot 10^{-3} \text{) unde,}$$

R^2 = determination coefficient (for training and test sets),

p-value = the probability of obtaining a significant model.

7.4. Conclusions

- Validation of the model according to the leave-one-out procedure;
- At the same time it was confirmed that with the decrease of the number of variables, the statistical significance of a model decreases.
- Based on the obtained model, it can be said that positions 18, 33 and 35 respectively are those whose statistical significance decreases (all these positions having a negative effect on the value of the logP coefficients).

CHAPTER 8. UNDERSTANDING THE EFFECT OF ARSENATES TREATMENT ON BREAST CANCER CELL LINES USING GENE EXPRESSION ANALYSIS

8.1. Materials and methods

In order to evaluate the effect of the arsenates ($2 \text{ HNO}_3 + \text{As}_2\text{O}_3 + 2 \text{ H}_2\text{O} \rightarrow 2 \text{ H}_3\text{AsO}_4 + \text{N}_2\text{O}_3$) on three breast cancer cell lines (double-positive MCF-7, triple negative Hs578T, negative triple MDA-MB-231) and a normal human mammary epithelial cell HUMEK for control, were investigated the changes to transcriptomic level, in particular, on modulation of apoptosis, autophagy and cell proliferation processes.

8.1.3. Analysing microarray data

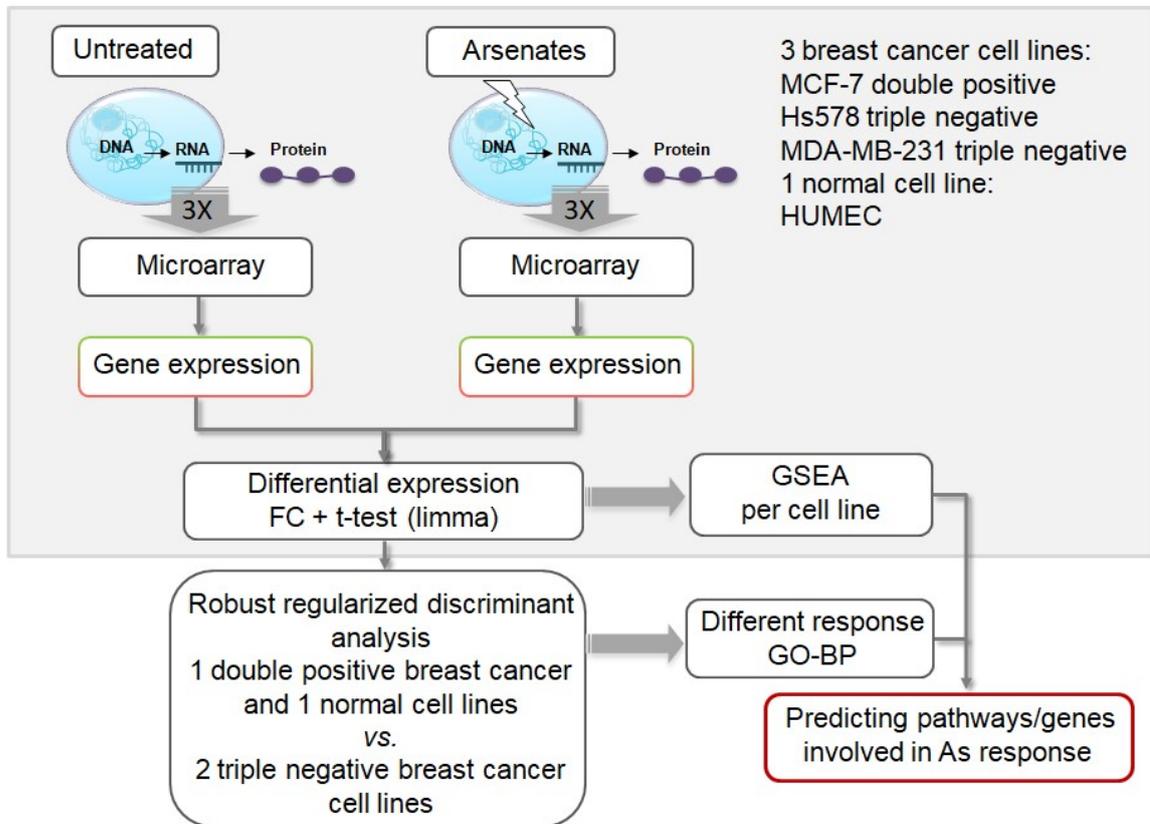


Figure 8.1. The transcriptomic analysis algorithm using microarray data

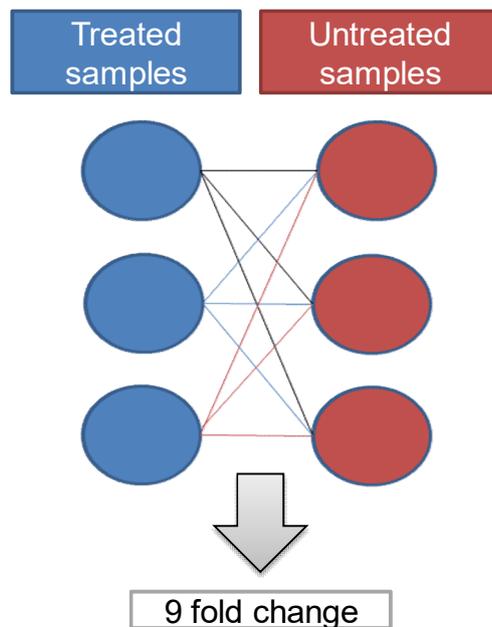


Figure 8.2. Fold change calculation

For each of the samples, treated and untreated, the level of gene expression (log₂ fold change) was calculated for each cell line resulting in a 9 fold change expression on each cell line (Figure 8.2).

Table 8.1. Prediction based on different parameters in RDA analysis

Alpha value	Delta value									
	0	0.33	0.7	1	1.333	1.667	2	2.333	2.667	3
0	2	2	1	1	5	7	18	18	18	18
0.11	0	2	10	10	12	10	11	18	18	18
0.22	0	0	9	10	11	11	10	11	18	18
0.33	0	0	9	10	10	12	11	10	11	11
0.44	0	0	9	9	10	11	12	11	10	11
0.55	0	0	2	9	10	10	10	11	12	11
0.66	0	0	0	5	9	9	10	10	10	11
0.77	0	0	0	0	3	9	9	9	10	10
0.88	0	0	0	0	0	0	0	1	3	7
0.99	0	0	0	0	0	0	0	0	0	0

In Table 8.1 we can find and observe all the data of the robust and calculated matrices by highlighting those values considered. The chosen value in orange is the size chosen in the RDA analysis.

8.2. Results and discussions

8.2.1. In vitro matrigel assay

The short-term in vitro matrigel culture test was performed in order to track the effects of arsenates on the cellular organization (Figure 8.3).

It is shown that in all cell lines, arsenates modulate processes involved in cellular elongation and reduce intercellular interactions and cellular cellular capacity. This effect is highlighted in triple negative Hs578T and MDA-MB-231 breast cancer cells (Figure 8.3).

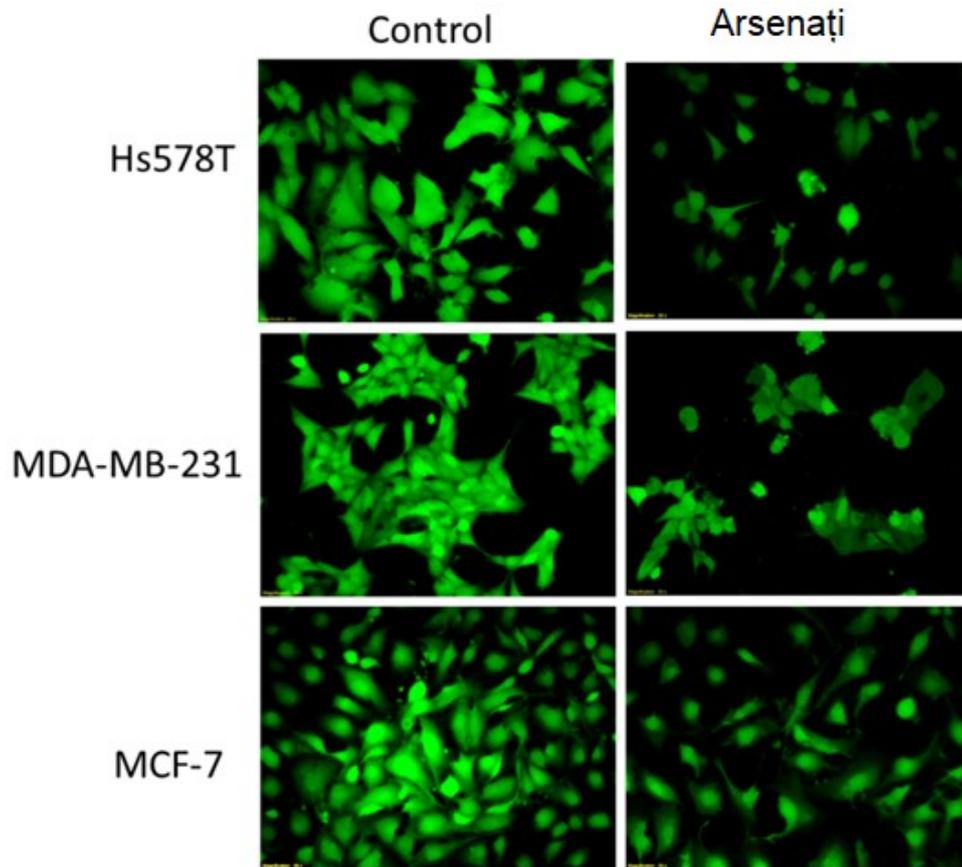


Figure 8.3. Short-term *in vitro* test on matrigel culture. Microscopic view of untreated cells (Control) vs. cells treated with arsenates.

8.2.2. Effect of arsenates on the regulation of autophagy and apoptosis

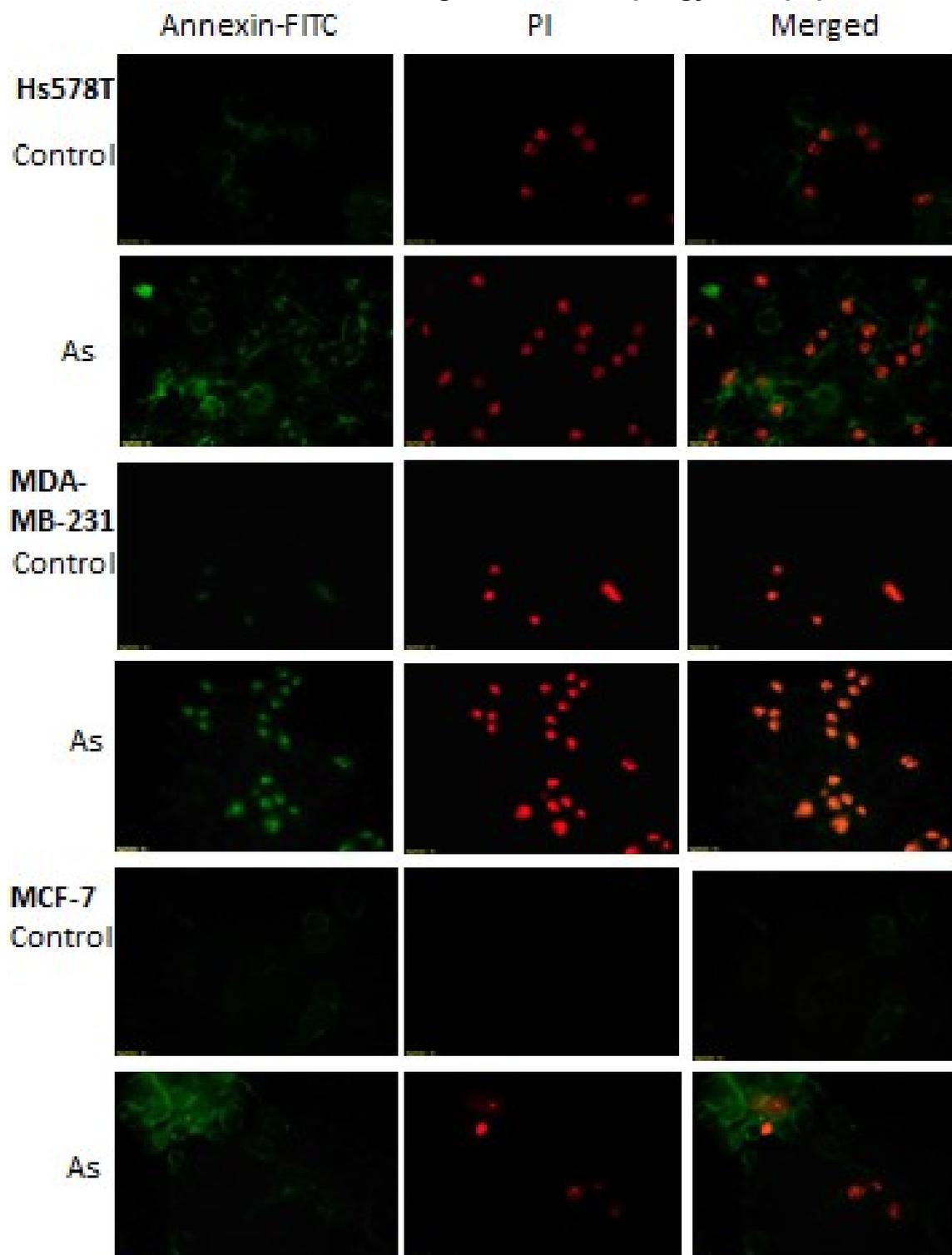


Figure 8.4.a. Evaluation of the apoptosis process by fluorescence microscopy (20X magnitude) as a result of cell exposure to arsenates (As - notation in the image) 50 nM.

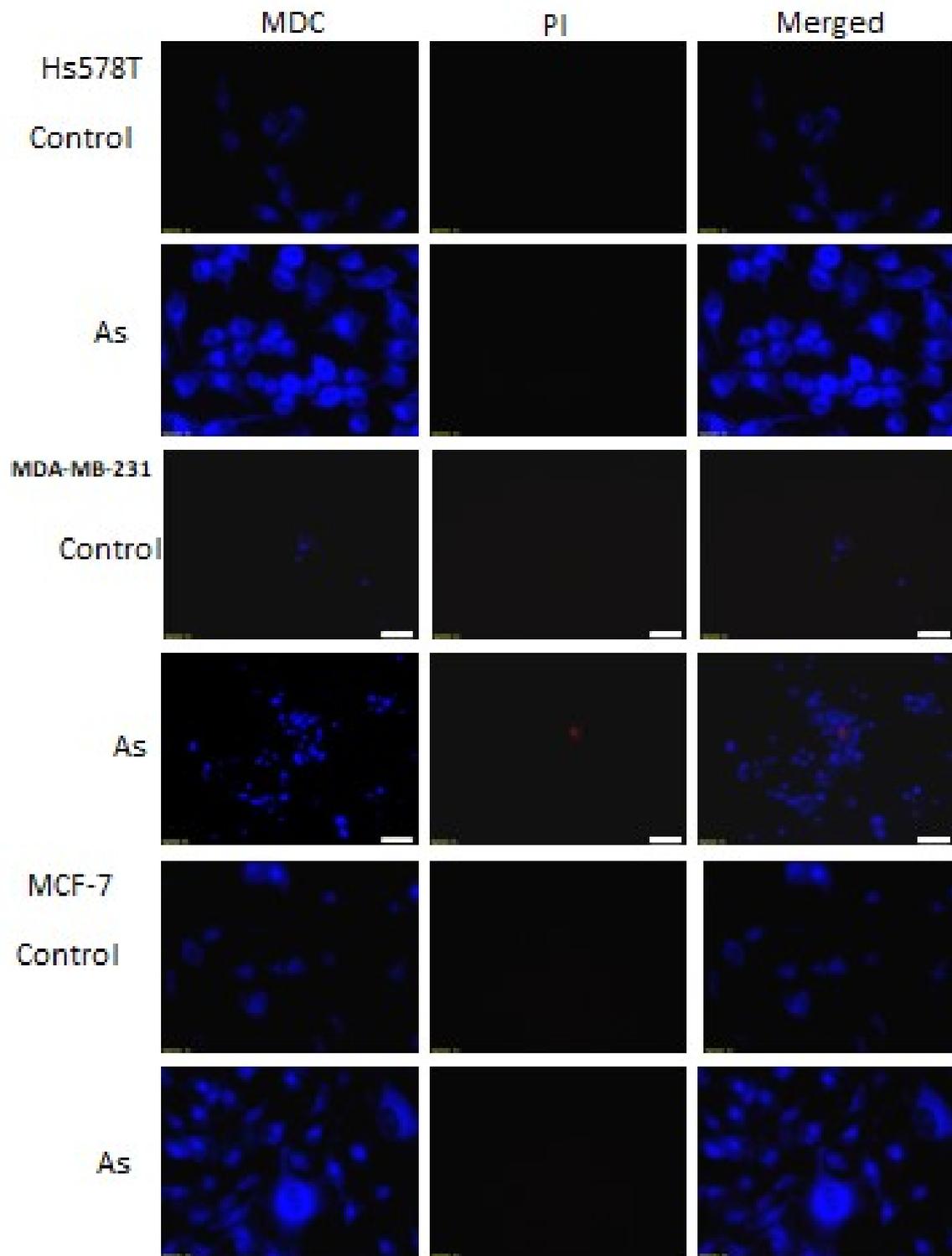


Figure 8.4.b. Evaluation of the autophagy process by fluorescence microscopy (20X magnitude) as a result of cell exposure to arsenacts (As-notation in the image) 50 nM

8.2.3. Shape of the data based on Pearson correlation

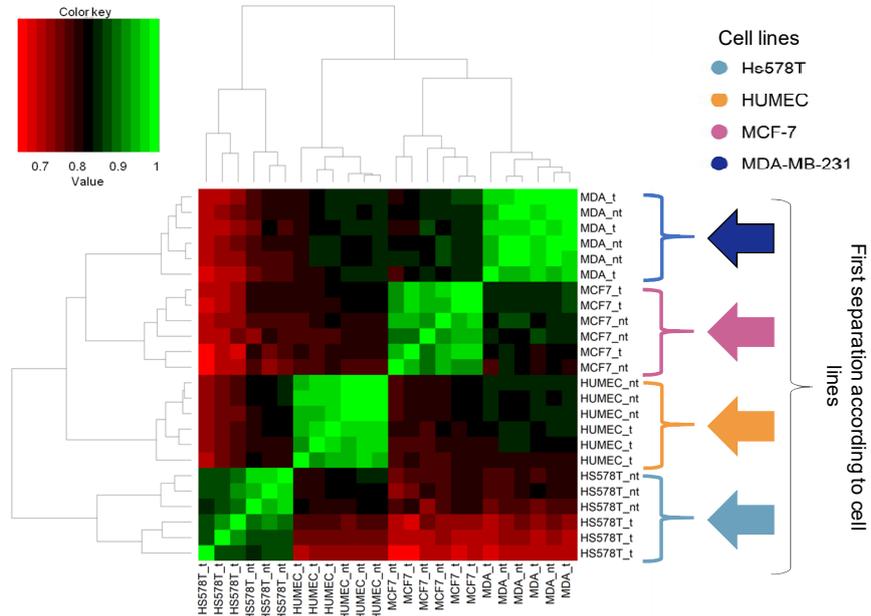


Figure 8.5. Arsenates modulate different responses in the four distinct cell lines

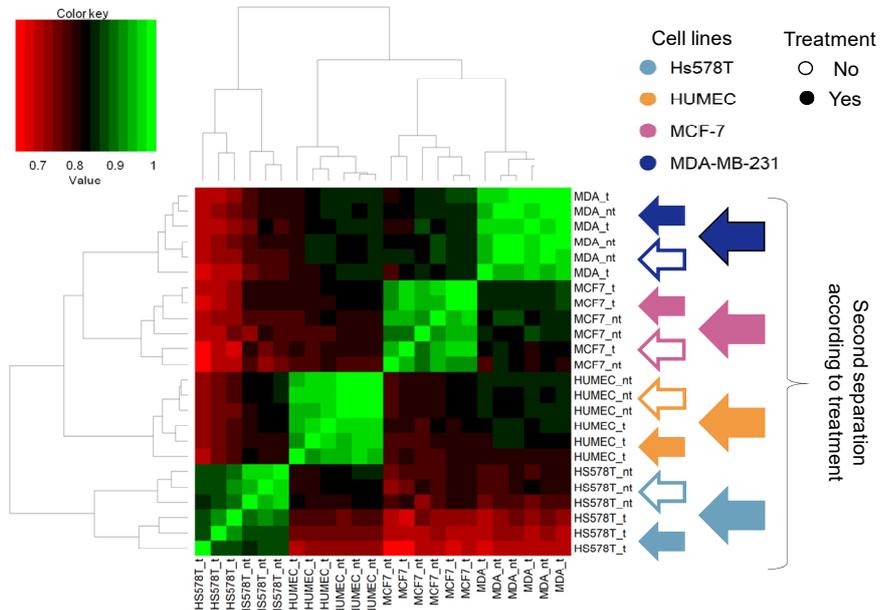


Figure 8.6. Different responses in the two conditions (treated-t, untreated-nt)

8.2.4. Principal component analysis

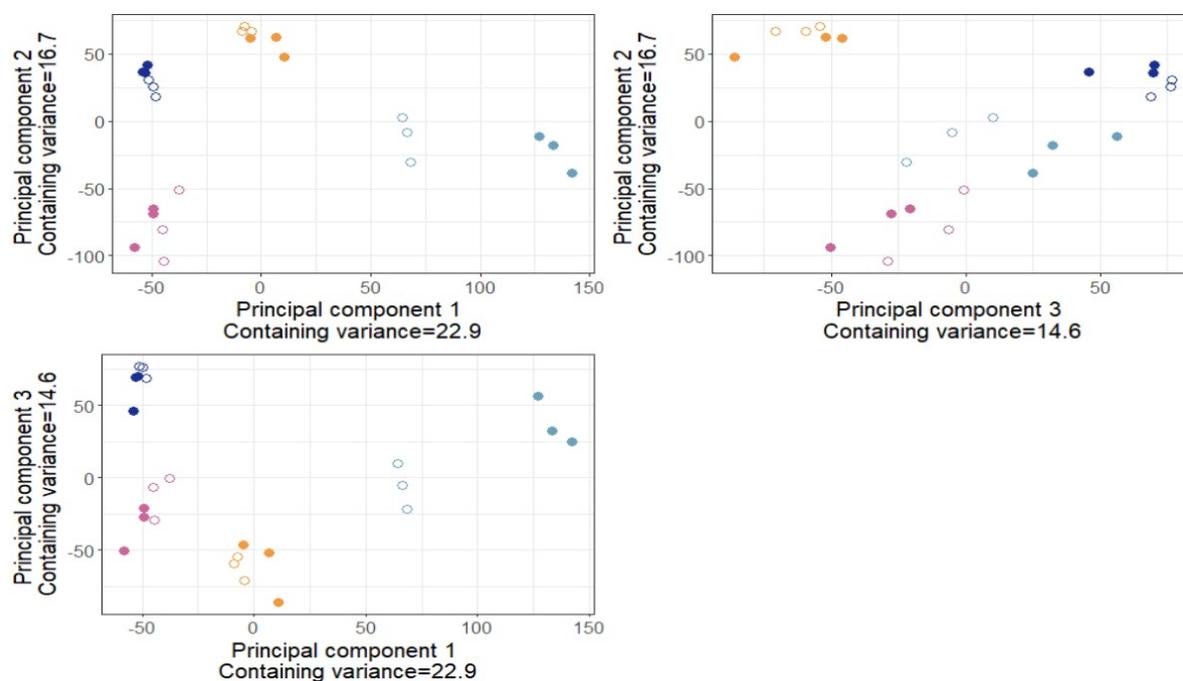


Figure 8.7. Each cell line is distinguished from a distinct area within the gene expression space, considering the three main components containing a variant of 22.9% (CP1), 16.7% (CP2), and 14.6% (CP3).

8.2.5. Gene set enrichment analysis

Table 8.2. Common enriched biological processes responding to arsenic in all four cell lines (FDR <0.15)

Name of biological process	FDR q-value	
	Hs578T/ MDA-MB-231	MCF-7/ HUMEC
DNA replication dependent nucleosome organization	0 0	0 0
Chromatin silencing at rDNA	0 0	0 0
Protein heterotetramerization	0 0	0.00018 0
Chromatin silencing	0 0.00064	0 0
Negative regulation of hematopoietic progenitor cell differentiation	0 0.01886	0.00027 0
Negative regulation of megakaryocyte	0	0

Name of biological process	FDR q-value	
	Hs578T/ MDA-MB-231	MCF-7/ HUMEC
differentiation	0.00284	0
Negative regulation of gene expression epigenetic	0 0.00083	0.00015 0
Positive regulation of gene expression epigenetic	0 0.04965	0.01009 0
Beta catenin TCF complex assembly	0 0.00326	0.03176 0
Protein heterooligomerization	0 0.04438	0.00748 0
Telomere organization	0 0.04186	0.00747 0
Gene silencing by RNA	0 0.01242	0.12874 0.00026
Regulation of gene silencing	0.00942 0.00100	0.00363 0.02088

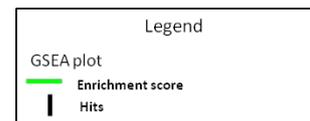
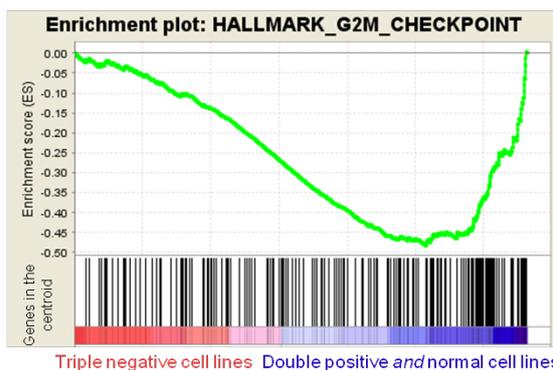
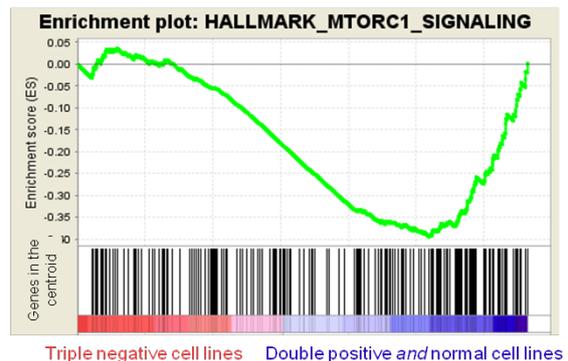
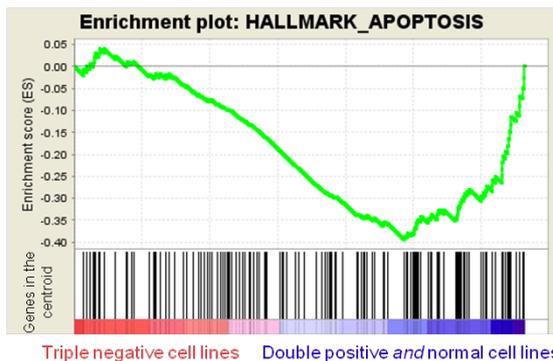


Figure 8.8. Genes are perturbed differently in mTOR apoptosis and G2M checkpoint signalling. ($q < 0.15$) On the x axis centroid values representing where are the genes more perturbed.

Data centroid values are the comparative response between triple negative (red) versus double negative and normal (blue) cell lines (Figure 8.8).

8.3. Conclusions

- As a global effect in the treatment of cancer cells with arsenates, it has been shown that by this the epigenetic regulation has been disturbed.
- Apoptosis and autophagy processes have been plagued in triple-negative breast cancer cell lines (Hs578T and MDA-MB-231).
- Biological pathways involving DNA repair mechanisms, chromatin organization and epigenetic regulation have been modulated to a high level by arsenates.
- Considering the dose of arsenates (50nM) to which the cells were subjected, and the effects on the above-mentioned biological processes, it can be concluded that the arsenates may be assigned to an anti-tumor alternative treatment for those tumors involving the disruption of the DNA repair process.
- The resulting transcriptomic data provides an insight into the effect of arsenates on processors and complex mechanisms involving inhibition or activation of tumor processes.

CHAPTER 9. UNDERSTANDING THERAPEUTIC EFFECTS OF LACTOBACILLUS ON INTESTINAL CELL LINES USING GENE EXPRESSION ANALYSIS

9.1. Materials and methods

9.1.4. Extraction of total RNA

Total RNA, both from lactobacilli-treated IPEC-1 cells and from untreated IPEC-1 cells, was isolated using the QiagenRNeasy midi kit (QIAGEN GmbH) following the supplier's recommended protocol Pistol et al., 2014). The quality and integrity of the samples were checked using the Agilent 2100 bioassay analyzer and the Agilent RNA 6000 nano kit (Agilent Technologies). The RIN (integrity number) score was found between 8-10. Purified RNA samples were stored at -80° C until use.

9.1.6. Statistical analysis of microarray data

All the genomic sequences of the *Sus Scrofa* were extrapolated to their human counterparts using Homology Based Annotation from the NCBI database (www.ncbi.nlm.nih.gov) and BLAST (Braicu et al., 2016).

Ingenuity Pathway (IPA; <http://www.ingenuity.com>) analysis was performed.

9.1.7. Validation of gene expression data

Validation of gene expression data was performed by quantitative Real-time PCR (RT-qPCR) analysis. The randomly selected four gene expression profiles (IL-1 β , TLR6, TLR4, IL-10) (Table 9.1) were measured by RT-qPCR in all samples considered according to the protocol shown in (Taranu et al., 2015).

Table 9.1. Validation of gene expression data considering the levels of expression of four genes (IL-1 β , TLR6, TLR4, IL-10) with geometric mean (Geomean) and Fold change (FC) after *Lactobacillus* treatment (LB).

Pathways	Genes	Geomean_Fold_LB	FC
Inflammatory response	TLR6	1.03	2.04
	TLR4	1.46	2.75
Cytokinesis	Il-1b	1.45	2.73
	IL10	1.8	3.48

9.2. Results and discussions

9.2.2. Functional classification of differentially expressed genes

The genes that were found to be differentially expressed after treatment were further subjected to a cluster assay and ranked in eight functional categories and signaling pathways as: signaling, cell signaling, proliferation, transcription factor, factors growth, cytokines, interleukins, inflammatory response (Table 9.2).

Table 9.2. The up- (red) and down- (green) regulated genes involved in the eight functional categories and signaling pathways

Pathways	Up regulated	Down regulated	Total
Transcription factor	597	60	657
Signaling	1735	76	1811

Cellular signaling	132	5	137
Proliferation	632	60	692
Cytokines	224	10	234
Interleukins	186	14	200
Inflammatory response	205	13	218
Growth factor	398	13	411

76 genes of 1811 were found to be down-regulated, significant gene suppression being observed for RGS2 genes with an expression level of -6.67 and p-value = 0.017 and OR1L8 with an expression level of -5.26 and p-value = 0.259 (Table 9.3).

Table 9.3. List of up- and down-regulated genes involved in cell signaling in IPEC-1 porcine epithelial cells

ID Gene	Gene symbol	Genes description	FC	Expression
Signaling				
GACC01000361	<i>nf1</i>	Neurofibromin 1 (NF1), transcript variant 1	10.20	Up
AK349266	<i>fuz</i>	Fuzzy homolog (Drosophila) (FUZ) Transcript variant 1	10.63 10.70	Up Up
XM-001928433	<i>or2m3</i>	Olfactory receptor, family 2, subfamily M, member 3 (OR2M3)	11.63	Up
AK345382	<i>entpd1</i>	Ectonucleoside triphosphate diphosphohydrolase 1 (ENTPD1), Transcript variant 1	12.13 13.45	Up Up
XM-003482962	<i>axin2</i>	AXIN 2 (axis inhibition protein)	13.93	Up
AB530146	<i>rgs2</i>	Regulator of G-protein signaling 2, 24kda	-6.67	down
XM-001925049	<i>or1l8</i>	Olfactory receptor, family 1, subfamily L, Member 8	-5.26 -4.17	down down
NM-001001861	<i>cxcl2</i>	Chemokine (C-X-C motif) ligand 2	-4.00	down
NM-214376	<i>areg</i>	Amphiregulin	-3.85	down
NM-214376	<i>areg</i>	Amphiregulin	-3.85	down
AY609724	<i>tcf21</i>	Transcription factor 21	-3.70	down

A significant number (692 genes) of differentially expressed genes involved in cell proliferation were observed, of which 632 were found to be over-expressed (of which 59 genes had an expression level of 4, e.g., NF1 with a 10.20 FC) and 60 sub-expressed genes such as AREG with an expression level of -3.85 FC, IL1a with an expression level of -3.33 FC (Table 9.4).

Table 9.4. List of up- and down-regulated genes involved in cell proliferation in IPEC-1 swine epithelial cells

ID Gene	Gene symbol	Genes description	FC	Expression
<u>Proliferation</u>				
GACC01000361	<i>nf1</i>	Neurofibromin 1 (NF1), transcript variant 1	10.20 10.70 13.45	Up
XM-003482962	<i>axin2</i>	AXIN 2 (axis inhibition protein)	13.93	Up
NM-214376	<i>areg</i>	Amphiregulin	-3.85 -3.45	down
NM-214029	<i>il1a</i>	Interleukin 1, alpha (IL1A)	-3.33	down
AY610314	<i>ube2v2</i>	Ubiquitin-conjugating enzyme E2 variant 2	-3.33	down

A group of genes (657) involved in transcription factors were found to be differentially expressed, 597 being over-expressed (with an expression level of 12.21 and p-value = 0.058 the TSHZ2, NF1 gene with a level of expression of 10.20 and p-value = 0.133, EMX1 with an expression level of 9.00 and p-value = 0.133) and 60 genes are sub-expressed (e.g., the PKNOX2 gene with an expression level of -8.33 and p-value = 0.015) (Table 9.5).

Table 9.5. List of up- and down-regulated genes involved in transcription factors in IPEC-1 porcine epithelial cells

ID Gene	Gene symbol	Genes description	FC	Expression
<u>Transcription factors</u>				
XM-003125031	<i>emx1</i>	EMX1 (empty spiracles homeobox 1)	9.00	Up
GACC01000361	<i>nf1</i>	Neurofibromin 1 (NF1), transcript variant 1	10.20	Up
AK347929	<i>tshz2</i>	TSHZ2 (teashirt zinc finger homeobox 2), transcript variant 1	12.21	Up
XM-003361490	<i>pknox2</i>	PBX/knotted 1 homeobox 2	-8.33	down

ID Gene	Gene symbol	Genes description	FC	Expression
AY609724	<i>tcf21</i>	Ref <i>Homo sapiens</i> transcription factor 21 (TCF21), transcript variant 2	-3.70	down

Table 9.6. List of up- and down-regulated genes involved in Inflammatory response in IPEC-1 porcine epithelial cells

ID Gene	Gene symbol	Genes description	FC	Expression
Inflammatory response				
XM-003131278	<i>prkca</i>	Protein kinaza C, alpha	5.13 5.58	Up
AK396677	<i>pla2g7</i>	Phospholipase A2, group VII (platelet-activating factor acetylhydrolase, plasma)	5.82	Up
XM-001929161	<i>osm</i>	Oncostatin M	6.19	Up
XM-003130465	<i>il20</i>	Interleukin 20	6.23	Up
AY669080	<i>bmp2</i>	Bone morphogenetic protein 2	6.87	Up
AK232615	<i>serpina3</i>	Serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3	9.85	Up
AK345252	<i>cxcl2</i>	Chemokine (C-X-C motif) ligand 2	-6.25	down
XM-003129107	<i>cxcl2</i>	Chemokine (C-X-C motif) ligand 2	-6.25	down
XM-003126166	<i>cxcl2</i>	Chemokine (C-X-C motif) ligand 2	-6.25	down
NM-001001861	<i>cxcl2</i>	Chemokine (C-X-C motif) ligand 2	-4.00	down
AY577905	<i>cxcl2</i>	Chemokine (C-X-C motif) ligand 2	-3.45	down
NM_214029	<i>il1a</i>	Interleukin 1, alpha	-3.33	down

9.2.3. Real-Time PCR validation

The microarray results were validated by qRT-PCR analysis for expression levels of four genes (IL-1 β , TLR6, TLR4, IL-10). Selected expression levels of the selected genes showed a near degree of over-expression in both microarray analysis (IL-1 β with 2.73 FC, TLR-6 with 2.04 FC, TLR-4 with 1.46 FC and IL-10 with 3.48 FC) and in the case of qRT-PCR analysis (IL-1 β with 2.32 FC, TLR-6 with 1.84 FC, TLR-4 with 1.74 FC and IL-10 with 3.77 FC) (Table 9.1). From both cases it can be concluded

that a good correlation between the results has been obtained. The method has the potential to eliminate the variability that could influence quantitatively the expression levels of the genes.

9.2.4. Pathway analysis

Among the canonical pathways involved are Wnt / β -catenin and the molecular mechanism involved in cancer development (Table 9.7), cellular functions, cell growth and proliferation, cell division, death (apoptosis) and cell survival (Table 9.8 and 9.9).

Table 9.7. Significantly expressed canonical pathways associated with treatment with lactobacilli following IPA analysis

Name	LB treatment	
	<i>p-value</i>	Ratio
The role of macrophages, fibroblasts and endothelial cells in rheumatoid arthritis	9.33E-21	56/309 (0.181)
The molecular mechanism of canc	4.29E-20	61/374 (0.163)
Wnt/ β -catenina	1.32E-18	39/169 (0.231)
Pluripotent human embryonic stem cells	1.33E-18	36/143 (0.252)

Table 9.8. Significantly expressed genes and associated cellular functions associated with lactobacilli treatment following IPA analysis

Name	LB treatment		
	<i>p-value</i>		Molecules
Cellular growth and proliferation	2.79E-20	-1.53E-124	598
Cellular Development	2.79E-20	-1.13E-117	572
Gene Expression	1.83E-33	-1.33E-117	439
Cellular Movement	3.35E-20	-2.04E-95	392
Cell death and Survival	4.95E-20	-5.24E-86	478

Table 9.9. Significantly expressed gene involved in affections and biological functions associated with lactobacilli treatment following IPA analysis

Name	LB treatment		
	<i>p-value</i>		Molecules
Cancer	2.41E-20	-1.11E-58	849
Organism Injury and Abnormalities	2.41E-20	-1.11E-58	860
Gastrointestinal Diseases	2.38E-20	-1.53E-42	761

Name	LB treatment		Molecules
	<i>p</i> -value		
Developmental Disorder	3.43E-21	-5.85E-42	282
Inflammatory Response	5.00E-33	-2.08E-43	362

Table 9.10. Number of significantly expressed genes involved in functional networks associated with lactobacilli treatment following IPA analysis

Rețele funcționale asociate	Score
Cell Signaling Cell-to-Cell Signaling and Interaction, Cell Cycle	52
Gene Expression, Cellular Development, Digestive System Development and Function	41
Gene Expression, Skeletal and Muscular Disorders, Skeletal and Muscular System Development and Function	37
Cellular Movement, Hematological System Development and Function, Immune Trafficking	37
Gene Expression, Hematological System Development and Function, Tissue Morphology	37

9.4. Conclusions

- The results obtained from the transcriptomic analysis indicate that the mixture of the three lactobacilli strains (*L. rhamnosus*, *L. plantarum* and *L. paracasei*) of concentration 1×10^8 CFU /mL differentially modulates gene expression, having a beneficial effect on functional epithelial barrier, on cell proliferation, inflammation and immune response (cytokines, chemokines) in IPEC-1 intestinal epithelial cells.
- The predominant effect of probiotics tested was gene activation, the genes involved in the signaling pathways being the most affected: 95% up- regulated. Many of the expressed genes are involved in cellular pathways and important biological functions. Compared to the control gene list (13950), 12678 over-expressed genes and 1272 down- expressions were found implicated in the predominant effect.

- Most of the modulated genes (1811) were associated with signaling pathways of which 121 up-regulated genes with a cut off of 2 and a reference expression level (FC) greater than 10. The lactobacilli mix had a significant effect on the pathway Wnt/ β -catenin signaling, for which the AXIN2 gene was found to be over-expressed with a fold change of 13.93, a β -catenin negative regulator that plays an important role in human cancer tumors.
- The results obtained from the microarray analysis highlight the effects of the lactobacilli mix on cell proliferation and transcription. The NF1 gene encoding the neurofibromin protein, a tumor suppressor that prevents uncontrolled cell proliferation, has an expression level greater than 10 FC.
- Induction of genes such as SERPINA 3, IL-20, OSM, GM-CSF, as well as suppression of CXCL-2 (MCP-1) and RGS2 genes and IL-18 proinflammatory cytokine highlights the protective role of lactobacilli in the epidermal barrier function inflammation and in activating the immune response.

CONCLUSIONS AND FUTURE PROSPECTS

- It has been shown that a set of inorganic compounds (Chapter 5) tend to form different equilibrium arrangements as a result of interaction with water molecules and in the absence of other ions (at an infinite dilution) the arrangements may have altered symmetry. At the same time, in solutions concentrated with 8 lithium and 4 potassium atoms, the formed dodecaded arrangements proved to be symmetrical and stable, whereas sodium dodecaders containing 6 atoms of sodium exhibit instability and altered symmetry.
- Regarding the similarity in classes of compounds with anti-inflammatory properties (Chapter 6), it has been shown that a series of natural compounds in the collected data set are similar to the structures of the drug compounds administered in the treatment of inflammatory diseases. Also, on the basis of

similarity, it has been shown that a natural compound and / or its derivatives can be found in a variety of other related plant extracts based on phylogeny.

- Following the behavior and arrangements of the organic molecules represented by the set of steroidal derivatives (Chapter 7), it was found that following their overlapping through the construction of the hypermolecule it was possible to estimate a structure-property relationship considering the mass of the fragments and the lipophilicity property . The relation between the structural characteristics and the considered property was found to be of high statistical significance, as shown by the validation of the model by the leave-one-out procedure, with a coefficient of determination $Q^2 = 0.9413$ after leave-one-out procedure and $R^2 = 0.961$ in estimated model.
- Following the treatment of arsenic cancer cells (Chapter 8), epigenetic regulation was shown to be disturbed, and apoptosis and autophagy processes were plagued in triple-negative breast cancer cell lines (Hs578T and MDA-MB-231) implicitly affirming that arsens act on complex processes and mechanisms that have the effect of inhibiting tumor processes. Also, following the treatment, the arsens have been shown to be involved in biological ways involving DNA repair mechanisms, chromatin organization and epigenetic regulation, modifying them to a high level.
- The probiotic solution mix (Chapter 9), applied as a treatment for porcine intestinal epithelial cells, differentiated gene expression by demonstrating beneficial effects on functional epithelial barrier, cell proliferation, inflammation and immune response.
- Understanding the structural properties and characterizing the molecular (in water and silico) behavior of both inorganic and organic compounds can lead to the explicability of their biological activities in vitro. On the basis of similarity, it may contribute to the selection and selection of the compounds to be tested in vitro for therapeutic purposes, depending on the condition for which treatment is desired.

References (Selection)

- Albrecht, B., Grant, G.H., Richards, W.G., 2004. Evaluation of structural similarity based on reduced dimensionality representations of protein structure. *Protein engineering, design & selection* 17, 425-432. doi:10.1093/protein/gzh049
- Alexander-Dann, B., Pruteanu, L.L., Oerton, E., Sharma, N., Berindan-Neagoe, I., Módos, D., Bender, A., 2018. Understanding and predicting compound-induced toxicity from transcriptomic data. *Molecular Omics*. doi:10.1039/c8mo00042e
- Allen, B.C.P., Grant, G.H., Richards, W.G., 2001. Similarity calculations using two-dimensional molecular representations. *Journal of Chemical Information and Computer Sciences* 41, 330-337. doi:10.1021/ci0003956
- Anjum, A., Jaggi, S., Varghese, E., Lall, S., Bhowmik, A., Rai, A., 2016. Identification of differentially expressed genes in RNA-seq data of *Arabidopsis thaliana*: A compound distribution approach. *Journal of Computational Biology* 23, 239-247. doi:10.1089/cmb.2015.0205
- Auer, P.L., Doerge, R.W., 2010. Statistical design and analysis of RNA sequencing data. *Genetics* 185, 405-416. doi:10.1534/genetics.110.114983
- Bender, A., 2005. Studies on molecular similarity. PhD thesis.
- Bender, A., Jenkins, J.L., Scheiber, J., Sukuru, S.C.K., Glick, M., Davies, J.W., 2009. How similar are similarity searching methods? A principal component analysis of molecular descriptor space. *Journal of Chemical Information and Modeling* 49, 108-119. doi:10.1021/ci800249s
- Bero, S.A., Muda, A.K., Choo, Y.H., Muda, N.A., Pratama, S.F., 2017. Similarity measure for molecular structure: A brief review. *Journal of Physics: Conference Series* 892, 1-8. doi:10.1088/1742-6596/892/1/012015
- Biella, C.D. a, Salvador, M.J., Dias, D.A., Dias-Baruffi, M., Pereira-Crott, L.S., 2008. Evaluation of immunomodulatory and anti-inflammatory effects and phytochemical screening of *Alternanthera tenella* Colla (Amaranthaceae) aqueous extracts. *Memorias do Instituto Oswaldo Cruz* 103, 569-577. doi:10.1590/S0074-02762008000600010
- Bleakley, K., Yamanishi, Y., 2009. Supervised prediction of drug-target interactions using bipartite local models. *Bioinformatics* 25, 2397-2403. doi:10.1093/bioinformatics/btp433
- Bose, B., 2013. Systems biology: A biologist's viewpoint. *Progress in Biophysics and Molecular Biology* 113, 358-368. doi:10.1016/j.pbiomolbio.2013.07.001
- Bourdon-Lacombe, J.A., Moffat, I.D., Deveau, M., Husain, M., Auerbach, S., Krewski, D., Thomas, R.S., Bushel, P.R., Williams, A., Yauk, C.L., 2015. Technical guide for applications of gene expression profiling in human health risk assessment of environmental chemicals. *Regulatory Toxicology and Pharmacology* 72, 292-309. doi:10.1016/j.yrtph.2015.04.010
- Braicu, C., Cojocneanu-Petric, R., Jurj, A., Gulei, D., Taranu, I., Gras, A.M., Marin, D.E., Berindan-Neagoe, I., 2016. Microarray based gene expression analysis of *Sus Scrofa* duodenum exposed to zearalenone: Significance to human health. *BMC Genomics* 17, 1-10. doi:10.1186/s12864-016-2984-8
- Breitling, R., Armengaud, P., Amtmann, A., Herzyk, P., 2004. Rank products: A simple, yet powerful, new method to detect differentially regulated genes in

- replicated microarray experiments. *FEBS Letters* 573, 83-92. doi:10.1016/j.febslet.2004.07.055
- Burnham, C.J., Petersen, M.K., Day, T.J.F., Iyengar, S.S., Voth, G.A., 2006. The properties of ion-water clusters. II. Solvation structures of Na⁺, Cl⁻, and H⁺ clusters as a function of temperature. *The Journal of Chemical Physics* 124, 24327-1-24327-9. doi:10.1063/1.2149375
- ChemBioDraw 14 . 0 User Guide, 2016.
- Chen, X., Yan, C.C., Luo, C., Ji, W., Zhang, Y., 2015. Constructing lncRNA functional similarity network based on lncRNA-disease associations and disease semantic similarity. *Nature Publishing Group* 1-12. doi:10.1038/srep11338
- Clegg, S.L., Brimblecombe, P., 1989. Solubility of ammonia in pure aqueous and multicomponent solutions. *The Journal of Physical Chemistry* 93, 7237-7248. doi:10.1021/j100357a041
- Çöltekin, A., Fabrikant, S.I., Lacayo, M., 2010. Exploring the efficiency of users' visual analytics strategies based on sequence analysis of eye movement recordings. *International Journal of Geographical Information Science* 24, 1559-1575. doi:10.1080/13658816.2010.511718
- Corney, D.C., Basturea, G.N., Jefferson, T., 2013. RNA-seq using next generation sequencing. *Materials and Methods* 3, 1-18. doi:dx.doi.org/10.13070/mm.en.3.203
- DeBlase, A.F., Wolke, C.T., Weddle, G.H., Archer, K.A., Jordan, K.D., Kelly, J.T., Tschumper, G.S., Hammer, N.I., Johnson, M.A., 2015. Water network-mediated, electron-induced proton transfer in [C₅H₅N·(H₂O)_n]⁺ clusters. *Journal of Chemical Physics* 143, 1-8. doi:10.1063/1.4931928
- Di Palma, T.M., Bende, A., 2013. Vacuum ultraviolet photoionization and ab initio investigations of methyl tert-butyl ether (MTBE) clusters and MTBE-water clusters. *Chemical Physics Letters* 561-562, 18-23. doi:10.1016/j.cplett.2013.01.025
- Ding, H., Takigawa, I., Mamitsuka, H., Zhu, S., 2013. Similarity-based machine learning methods for predicting drug-target interactions: A brief review. *Briefings in Bioinformatics* 15, 734-747. doi:10.1093/bib/bbt056
- Diudea, M.V., Vizitiu, A.E., Janezic, D., 2007. Cluj and related polynomials applied in correlating studies. *Journal of Chemical Information and Modeling* 47, 864-874. doi:10.1021/ci600482j
- Diudea, M. V., Gutman, I., Jäntschi, L., 2001. *Molecular topology*, Nova Science Publishers. doi:10.1007/BF01151278
- Dodge, S., 2011. Exploring movement using similarity analysis.
- Dong, J., Cao, D.S., Miao, H.Y., Liu, S., Deng, B.C., Yun, Y.H., Wang, N.N., Lu, A.P., Zeng, W. Bin, Chen, A.F., 2015. ChemDes: An integrated web-based platform for molecular descriptor and fingerprint computation. *Journal of Cheminformatics* 7, 1-10. doi:10.1186/s13321-015-0109-z
- Doucet, J.-P., Weber, J., 1996. Molecular similarity. *Computer-Aided Molecular Design* 328-362. doi:10.2174/1568026043451186
- Egieyeh, S., Syce, J., Christoffels, A., Malan, S.F., 2016. Exploration of scaffolds from natural products with antiplasmodial activities, currently registered antimalarial drugs and public malarial screen data. *Molecules* 21, 1-14. doi:10.3390/molecules21010104

- Fournier, R., 2008. Geometric structure of the magic number clusters $\text{Li}_n\text{Na}_{8-n}$, $\text{Na}_n\text{K}_{8-n}$, and $\text{K}_n\text{Li}_{8-n}$. *Journal of Computational Methods in Science and Engineering* 8, 331-342.
- Galashev, A.E., 2013. A computer study of ammonium adsorption on water clusters. *Russian Journal of Physical Chemistry B* 7. doi:10.1134/S1990793113050047
- Grayson, N.E., Taormina, A., Twarock, R., 2009. DNA duplex cage structures with icosahedral symmetry. *Theoretical Computer Science* 410, 1440-1447. doi:10.1016/j.tcs.2008.12.005
- Guerra, D., David, J., Restrepo, A., 2014. Hydrogen bonding in the binary water/ammonia complex. *Journal of Computational Methods in Sciences and Engineering* 14, 93-102. doi:10.3233/JCM-140487
- Hardcastle, T.J., Kelly, K.A., 2010. baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics* 11, 422. doi:10.1186/1471-2105-11-422
- Harsa, A.M., Harsa, T.E., Bolboacă, S.D., Diudea, M. V, 2014. QSAR in flavonoids by similarity cluster prediction. *Current Computer-Aided Drug Design* 10, 115-128. doi:1875-6697/14 \$58.00+.00
- Hettne, K.M., Boorsma, A., van Dartel, D.A., Goeman, J.J., de Jong, E., Piersma, A.H., Stierum, R.H., Kleinjans, J.C., Kors, J.A., 2013. Next-generation text-mining mediated generation of chemical response-specific gene sets for interpretation of gene expression data. *BMC medical genomics* 6, 2. doi:10.1186/1755-8794-6-2
- Hong, F., Breitling, R., McEntee, C.W., Wittner, B.S., Nemhauser, J.L., Chory, J., 2006. RankProd: A bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics* 22, 2825-2827. doi:10.1093/bioinformatics/btl476
- Hvelplund, P., Kurtén, T., Støchkel, K., Ryding, M.J., Nielsen, S.B., Uggerud, E., 2010. Stability and structure of protonated clusters of ammonia and water, $\text{H}+(\text{NH}_3)_m (\text{H}_2\text{O})_n$. *The Journal of Physical Chemistry A* 114, 7301-7310. doi:10.1021/jp104162k
- Janeiro-Barral, P.E., Mella, M., 2006. Study of the structure, energetics, and vibrational properties of small ammonia clusters $(\text{NH}_3)_n$ ($n = 2-5$) using correlated ab initio methods. *Journal of Physical Chemistry A* 110, 11244-11251. doi:10.1021/jp063252g
- Jäntschi, L., 2000. Predicția proprietăților fizico-chimice și biologice cu ajutorul descriptorilor matematici. *Universitatea „Babeș-Bolyai” din Cluj-Napoca*.
- Jäntschi, L., Pruteanu, L.L., Cozma, A.C., Bolboacă, S.D., 2015. Inside of the linear relation between dependent and independent variables. *Computational and Mathematical Methods in Medicine* 2015, 1-11. doi:10.1155/2015/360752
- Jiang, Y., Gerhold, D.L., Holder, D.J., Figueroa, D.J., Bailey, W.J., Guan, P., Skopek, T.R., Sistare, F.D., Sina, J.F., 2007. Diagnosis of drug-induced renal tubular toxicity using global gene expression profiles. *Journal of Translational Medicine* 5, 47. doi:10.1186/1479-5876-5-47
- Kadota, K., Nakai, Y., Shimizu, K., 2008. A weighted average difference method for detecting differentially expressed genes from microarray data. *Algorithms for Molecular Biology* 3, 1-12. doi:10.1186/1748-7188-3-8
- Kochev, N., Monev, V., Bangov, I., 2003. 6. Searching chemical structures. *Chemoinformatics: A Textbook* 291-318. doi:10.1021/c160012a009

- Kukurba, K.R., Montgomery, S.B., 2016. RNA sequencing and analysis. *Cold Spring Harbor Protocol* 2015, 951-969. doi:10.1101/pdb.top084970.RNA
- Laage, D., Hynes, J.T., 2007. Reorientational dynamics of water molecules in anionic hydration shells. *Proceedings of the National Academy of Sciences* 104, 11167-11172. doi:10.1073/pnas.0701699104
- Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.-P., Subramanian, A., Ross, K.N., Reich, M., Hieronymus, H., Wei, G., Armstrong, S., Haggarty, S.J., Clemons, P., Wei, R., Carr, S., 2006. The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313, 1929-1935. doi:10.1126/science.1132939
- Legendre, A.M., 1805. Nouvelles méthodes pour la détermination des orbites des comètes [New methods for the determination of the orbits of comets] (in French). Paris: F. Didot.
- Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15, 550. doi:10.1186/s13059-014-0550-8
- Ma, C., Wang, L., Xie, X.Q., 2011. GPU accelerated chemical similarity calculation for compound library comparison. *Journal of Chemical Information and Modeling* 51, 1521-1527. doi:10.1021/ci1004948
- Milne, G.W.A., 2010. Computer software review of ChemBioDraw 12.0. *Journal of Chemical Information and Modeling*. doi:10.1021/ci100385n
- Minkin, V.I., 1999. Glossary of terms used in theoretical organic chemistry (IUPAC Recommendations 1999). *Pure and Applied Chemistry* 71, 1919-1981. doi:10.1351/pac199971101919
- Minkowski, H., 1953. *Geometrie der Zahlen*. Chelsea Pub. Co., New York.
- Mishra, P., Tripathi, V., Yadav, B.S., 2010. In silico QSAR modeling and drug development process. *GERF Bulletin of Biosciences* 1, 37-40.
- Nidhi, Glick, M., Davies, J.W., Jenkins, J.L., 2006. Prediction of biological targets for compounds using multiple-category bayesian models trained on chemogenomics databases. *Journal of Chemical Information and Modeling* 46, 1124-1133. doi:10.1021/ci060003g
- Orozcoa, A.A., Audouzeb, K., Brunaka, S., Tabourea, O., 2016. In silico systems pharmacology to assess drug's therapeutic and toxic effects. *Bentham Science Publishers* 0, 1-10. doi:10.1174/1381611811666160907093115
- Pearson, K., 1895. On the mathematical theory of evolution. VII. Mathematical Contributions to the theory of evolution— III. Regression, heredity, and Panmixia. *Philosophical Transactions of the Royal Society of London* 253-318.
- Pistol, G.C., Gras, M.A., Marin, D.E., Israel-Roming, F., Stancu, M., Taranu, I., 2014. Natural feed contaminant zearalenone decreases the expressions of important pro-and anti-inflammatory mediators and mitogen-activated protein kinase/NF-kB signalling molecules in pigs. *British Journal of Nutrition* 111, 452-464. doi:10.1017/S0007114513002675
- Pruteanu, L., Ersali, S., Bolboacă, S., 2016. A QSPR model for steroids. *Studia Universitatis Babeş-Bolyai Chemia* 61, 71-78.
- Pruteanu, L.L., Jäntschi, L., Ungureşan, M.L., Bolboacă, S.D., 2016. Models of monovalent ions dissolved in water. *Studia Universitatis Babeş-Bolyai Chemia* 61, 151-162.

- Rarey, M., Dixon, J.S., 1998. Feature trees: a new molecular similarity measure based on tree matching. *Journal of computer-aided molecular design* 12, 471-490. doi:10.1023/A:1008068904628
- Reynolds, C.A., Burt, C., Graham Richards, W., 1992. A linear molecular similarity index. *Quantitative Structure-Activity Relationships* 11, 34-35. doi:10.1002/qsar.19920110106
- Rininger, J.A., DiPippo, V.A., Gould-Rothberg, B.E., 2000. Differential gene expression technologies for identifying surrogate markers of drug efficacy and toxicity. *Drug Discovery Today* 5, 560-568. doi:10.1016/S1359-6446(00)01597-X
- Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2009. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139-140. doi:10.1093/bioinformatics/btp616
- Rogers, J., Tanimoto, T.T., 1960. A computer program for classifying plants. *Science* 132, 1115-1118. doi:10.1126/science.132.3434.1115
- RStudio, 2014. Markdown. Rstudio 4-5. doi:10.1002/wics.1348
- Russell, J.D., 2006. NIST Computational chemistry comparison and benchmark database. NIST standard reference database number 101.
- Ryding, M.J., Ruusuvoori, K., Andersson, P.U., Zatul, A.S., McGrath, M.J., Kurtén, T., Ortega, I.K., Vehkamäki, H., Uggerud, E., 2012. Structural rearrangements and magic numbers in reactions between pyridine-containing water clusters and ammonia. *Journal of Physical Chemistry A* 116, 4902-4908. doi:10.1021/jp3021326
- Sander, T., Freyss, J., Kor, M. Von, Rufener, C., 2015. DataWarrior: An open-source program for chemistry aware data visualization and analysis. *Journal of Chemical Information and Modeling* 55, 460-473. doi:10.1021/ci500588j
- Shi, W., Bugrim, A., Nikolsky, Y., Nikolskya, T., Brennan, R.J., 2008. Characteristics of genomic signatures derived using univariate methods and mechanistically anchored functional descriptors for predicting drug- and xenobiotic-induced nephrotoxicity. *Toxicology Mechanisms and Methods* 18, 267-276. doi:10.1080/15376510701857072
- Singh, J., Singh, S., Shaik, B., Deeb, O., Sohani, N., Agrawal, V.K., Khadikar, P. V., 2008. Mutagenicity of nitrated polycyclic aromatic hydrocarbons: A QSAR investigation. *Chemical Biology and Drug Design* 71, 230-243. doi:10.1111/j.1747-0285.2008.00629.x
- Sirci, F., Napolitano, F., Pisonero-Vaquero, S., Carrella, D., Medina, D.L., di Bernardo, D., 2017. Comparing structural and transcriptional drug networks reveals signatures of drug activity and toxicity in transcriptional responses. *npj Systems Biology and Applications* 3, 23. doi:10.1038/s41540-017-0022-3
- Stears, L., Martinskt, T., Schena, M., 2003. Trends in microarray analysis. *Nature Medicine* 9, 140-145.
- Taranu, I., Braicu, C., Marin, D.E., Pistol, G.C., Motiu, M., Balacescu, L., Beridan Neagoe, I., Burlacu, R., 2015. Exposure to zearalenone mycotoxin alters in vitro porcine intestinal epithelial cells by differential gene expression. *Toxicology Letters* 232, 310-325. doi:10.1016/j.toxlet.2014.10.022
- Taranu, I., Marin, D.E., Braicu, C., Pistol, G.C., Sorescu, I., Pruteanu, L., Beridan-Neagoe, I., Vodnar, D.C., 2018. In vitro transcriptome response to Probiotic Lactobacilli mixture in Intestinal Porcine Epithelial Cell line. *International Journal*

- of Molecular Sciences 19, 1-16.
- Tarca, A.L., Romero, R., Draghici, S., 2006. Analysis of microarray experiments of gene expression profiling. *American Journal of Obstetrics and Gynecology* 195, 373-388. doi:10.1016/j.ajog.2006.07.001
- Taylor, R.C., Acquah-Mensah, G., Singhal, M., Malhotra, D., Biswal, S., 2008. Network inference algorithms elucidate Nrf2 regulation of mouse lung oxidative stress. *PLoS Computational Biology* 4. doi:10.1371/journal.pcbi.1000166
- Terfloth, L., 2003. *Chemoinformatics: A Textbook*. 8. Calculation of structure descriptors. Wiley-VCH Verlag.
- Thomson, W., Kelvin, Lord, Nikolova, N., Jaworska, J., 2003. Approaches to measure chemical similarity - a Review. *QSAR & Combinatorial Science* 22, 1006-1026. doi:10.1002/qsar.200330831
- Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., Pachter, L., 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature biotechnology* 31, 46-53. doi:10.1038/nbt.2450
- Trapnell, C., Williams, B. a, Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., Pachter, L., 2011. Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. *Nature Biotechnology* 28, 511-515. doi:10.1038/nbt.1621. Transcript
- Velkoborsky, J., Hoksza, D., 2016. Scaffold analysis of PubChem database as background for hierarchical scaffold - based visualization. *Journal of Cheminformatics* 8, 1-14. doi:10.1186/s13321-016-0186-7
- Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10, 57-63. doi:doi:10.1038/nrg2484
- Willett, P., 2011. Similarity-based data mining in files of two-dimensional chemical structures using fingerprint measures of molecular resemblance. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 1, 241-251. doi:10.1002/widm.26
- Willett, P., Barnard, J.M., Downs, G.M., 1998. Chemical similarity searching. *Journal of Chemical Information and Computer Sciences* 38, 983-996. doi:10.1021/ci9800211
- Yona, G., Dirks, W., Rahman, S., Lin, D.M., 2006. Effective similarity measures for expression profiles. *Bioinformatics* 22, 1616-1622. doi:10.1093/bioinformatics/btl127
- Yu, Z.D., Bawden, C.S., Henderson, H.V., Nixon, A.J., Gordon, S.W., Pearson, A.J., 2006. Microarrays as a discovery tool in wool genomics. *Proceedings of the New Zealand Society of Animal Production* 66, 129-133. doi:10.1079/BJN19660078
- Zhang, B., Horvath, S., 2005. A general framework for weighted gene Co-Expression network analysis. *Statistical applications in genetics and molecular biology* 4. doi:10.2202/1544-6115.1128

Lista publicațiilor

- Jäntschi, L., **Pruteanu**, L.L., Cozma, A.C., Bolboacă, S.D., 2015. Inside of the linear relation between dependent and independent variables. **Computational and Mathematical Methods in Medicine** 2015, 1-11. doi:10.1155/2015/360752.
- Pruteanu**, L.L., Jäntschi, L., Ungureșan, M.L., Bolboacă, S.D., 2016. Models of monovalent ions dissolved in water. **Studia Universitatis Babes-Bolyai Chemia** 61, 151-162.
- Pruteanu**, L.L., Ersali, S., Bolboacă, S.D., 2016. A QSPR model for steroids. **Studia Universitatis Babes-Bolyai Chemia** 61, 71-78.
- Ungureșan, M.L., **Pruteanu**, L.L., Jäntschi, L., 2016. Correlation study among boiling temperature and heat of vaporization. **Studia Universitatis Babes-Bolyai Chemia** 61, 223-234.
- Jäntschi, L., Bálint, D., **Pruteanu**, L.L., Bolboacă, S.D., 2016. Elemental factorial study on one-cage pentagonal face nanostructure congeners. **Materials Discovery** 5, 14-21. doi:10.1016/j.md.2016.12.001.
- Pruteanu**, L.L., Jäntschi, L., Crăciunaș, C., 2017. Molecular Level Investigation of Staphylococci's Resistance Mechanisms to Antibiotics. **Notulae Scientia Biologicae** 9, 307. doi:10.15835/nsb9310153.
- Alexander-Dann, B., **Pruteanu**, L.L., Oerton, E., Sharma, N., Berindan-Neagoe, I., Módos, D. and Bender, A., 2018. Understanding and predicting compound-induced toxicity from transcriptomic data. **Molecular Omics**. doi:10.1039/c8mo00042e - in press.
- Taranu, I., Marin, D.E., Braicu, C., Pistol, G.C., Sorescu, I., **Pruteanu**, L., Berindan-Neagoe, I., Vodnar, D.C., 2018. *In vitro* transcriptome response to Probiotic *Lactobacilli* mixture in Intestinal Porcine Epithelial Cell line. **International Journal of Molecular Sciences**, 19. - acceptat.